



**BNL-72286-2006**

## ***Computational Science Center***

***in affiliation with the Departments of***

**Applied Mathematics and Statistics and  
Computer Science  
Stony Brook University**

**[www.bnl.gov/csc](http://www.bnl.gov/csc)**

**November 2006**

**Brookhaven National Laboratory**

**P.O. Box 5000  
Upton, NY 11973-5000  
[www.bnl.gov](http://www.bnl.gov)**

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.



# BROOKHAVEN COMPUTATIONAL SCIENCE CENTER

## Page

1.	Computational Science @ Brookhaven.....	3
2.	Overview.....	3
3.	Personnel.....	4
4.	Research.....	7
4.1	SciDAC at CSC.....	8
4.1.1	ITAPS.....	8
4.1.2	TOPS.....	8
4.1.3	Advanced Computing for 21 <sup>st</sup> Century Accelerator Science & Technology.....	9
4.1.4	Statistical Approaches to Aerosol Dynamics for Climate Simulation.....	9
4.2	Biological and Environmental Research.....	10
4.2.1	Binding of Anti-Fusion Peptides with HIVgp41.....	10
4.2.2	Molecular Dynamics Study of HIV-Protease.....	12
4.2.3	MD Simulations of the Activation of the Adenovirus Proteinase.....	13
4.2.4	MD Simulations of Biotoxins and of PDZ Domains.....	15
4.2.5	Visualization and Data Mining for the Detection of Cancer.....	17
4.2.6	Atmospheric Aerosols.....	20
4.3	Advanced Scientific Computing.....	22
4.3.1	ITAPS: Atomization and Spray.....	22
4.3.2	ITAPS: Frontier-Lite.....	24
4.3.3	Magnetohydrodynamics of Free Surface Flows.....	27
4.3.4	Pellet Ablation in Tokamak Refueling Process.....	29
4.3.5	Linear Algebra and QCDOC on Top 500.....	30
4.3.6	Molecular Dynamics on the QCDOC.....	31
4.3.7	Highly Scalable FFT and Molecular Dynamics Algorithms for QCDOC.....	31
4.3.8	Electrodynamics on QCDOC and BG/L.....	34
4.3.9	Parallel Heisenberg Spin Model on Supercomputer Architectures.....	35
4.3.10	Comparison of Two Methods for Elliptic Boundary Value Problems.....	37
4.3.11	Uncertainty Quantification.....	39
4.3.12	ViStA: Visual Statistical Analyzer.....	41
4.3.13	Parallel Visualization of Large Data Sets.....	43
4.3.14	Visualization Collaboratory.....	45
4.3.15	Cluster Computing.....	48
4.4	Basic Energy Sciences.....	48
4.4.1	Scalable Localizable Density Functional Theory.....	48
4.4.2	First Principles Calculations of Zinc-Oxide NanoWires.....	50
4.4.3	Computational Nanomagnets.....	51
4.4.4	DFT Study of Isocyanide Adsorption on Gold (111) Surface.....	52
4.4.5	Direct Numerical Simulation of Multiphase Flows with Phase Transitions.....	54
4.5	High Energy and Nuclear Physics.....	56
4.5.1	Neutrino Factory/Muon Collider Target.....	56
4.5.2	Modeling of Wake Fields and Impedances in Accelerators.....	57
4.5.3	Unified Accelerator Library: SIMBAD.....	59
4.5.4	Spallation Neutron Source.....	61
4.5.5	Grid Computing: MonALISA.....	63



## **1. COMPUTATIONAL SCIENCE @ BROOKHAVEN**

Computational Science is an integral component of Brookhaven's multi science mission, and is a reflection of the increased role of computation across all of science.

Brookhaven currently has major efforts in data storage and analysis for the Relativistic Heavy Ion Collider (RHIC) and the ATLAS detector at CERN, and in quantum chromodynamics. The Laboratory is host for the QCDOC machines (quantum chromodynamics on a chip), 10 teraflop/s computers which boast 12,288 processors each. There are two here, one for the Riken/BNL Research Center and the other supported by DOE for the US Lattice Gauge Community and other scientific users.

A 100 teraflop/s supercomputer will be installed at Brookhaven in the coming year, managed jointly by Brookhaven and Stony Brook, and funded by a grant from New York State. This machine will be used for computational science across Brookhaven's entire research program, and also by researchers at Stony Brook and across New York State. With Stony Brook, Brookhaven has formed the New York Center for Computational Science (NYCCS) as a focal point for interdisciplinary computational science, which is closely linked to Brookhaven's Computational Science Center (CSC).

## **2. OVERVIEW**

The CSC has established a strong program in computational science, with an emphasis on nanoscale electronic structure and molecular dynamics, accelerator design, computational fluid dynamics, medical imaging, parallel computing and numerical algorithms. We have been an active participant in DOE's SciDAC program (Scientific Discovery through Advanced Computing). We are also planning a major expansion in computational biology in keeping with Laboratory initiatives.

Additional laboratory initiatives with a dependence on a high level of computation include the development of hydrodynamics models for the interpretation of RHIC data, computational models for the atmospheric transport of aerosols, and models for combustion and for energy utilization.

The CSC was formed to bring together researchers in these areas and to provide a focal point for the development of computational expertise at the Laboratory. These efforts will connect to and support the Department of Energy's long range plans to provide Leadership class computing to researchers throughout the Nation.

Recruitment for six new positions at Stony Brook to strengthen its computational science programs is underway. We expect some of these to be held jointly with BNL.

### 3. PERSONNEL

The Director of the Computational Science Center (CSC) is James Davenport; James Glimm is Head of Applied Mathematics; and the Administrative Assistant is Claire Lamberti. CSC is located in Building 463 (30 Bell Avenue), Room 255, at Brookhaven National Laboratory, Upton NY 11973-5000. The fax number for CSC is 631-344-5751.

CSC grew out of the Center for Data Intensive Computing, which was established in 1999. It currently has a staff of 16. As part of an active, ongoing collaboration with graduate students from Stony Brook, over thirty Ph.D. students and several undergraduates hold scientific positions in CSC.

TABLE 1. CSC STAFF		
Name	Position	Phone/email
James Davenport	Director	631-344-3789 jdaven@bnl.gov
James Glimm	Head, Applied Mathematics	631-344-8155 glimm@bnl.gov
David Keyes	Visiting Scientist	631-344-5752 kd2112@columbia.edu
Nicholas D'Imperio	Physics Associate	631-344-8607/8589 dimperio@bnl.gov
Stratos Efstathiadis	Technology Architect	631-344-6080 stratos@bnl.gov
Mark Hybertsen	Physicist	631-344-5996 mhyberts@bnl.gov
Michael McGuigan	Computer Scientist	631-344-2695 mcguigan@bnl.gov
Roman Samulyak	Computational Scientist	631-344-3304 rosamu@bnl.gov
Leonard Slate	Advanced Technology Engineer	631-344-4102 slatest@bnl.gov
David Stampf	Senior Technology Architect	631-344-4148 drs@bnl.gov
Robert Bennett	Research Associate	631-344-2089 robertb@bnl.gov
Kab Seok Kang	Research Associate	631-344-3580 kskang@bnl.gov
Tianshi Lu	Research Associate	631-344-5752 tlu@bnl.gov
Claire Lamberti	Administrative Assistant	631-344-3051 lamberti@bnl.gov

**TABLE 2. CSC AFFILIATES**

<b>Name</b>	<b>Affiliation</b>	<b>Expertise</b>
Philip Allen	SB/Physics	Condensed Matter Theory
Carl Anderson	BNL/Biology	Protein Structures
Carmen Benkovitz	BNL/Environmental Sciences	Atmospheric Transport
Helene Benveniste	BNL/Directorate	Neuroscience
Natasa Bozovic	San Jose State University	Computational Science
Michael Creutz	BNL/Physics	Lattice Gauge Theory
Yuefan Deng	SB/Applied Math. & Statistics	Molecular Dynamics
Erwin George	SB/Applied Math & Statistics	Fluid Dynamics
David Green	SB/Applied Math & Statistics	Computational Biology
John Grove	Los Alamos National Lab.	Computational Science
Jiansheng Jiang	BNL/Biology	Protein Structures
Chulwoo Jung	BNL/Physics	Lattice QCD
Arie Kaufman	SB/Computer Science	Visualization
Richard Kenway	UK QCDOC	Lattice QCD
Jerome Lauret	BNL/Physics	Grid Computing
Xiaolin Li	SB/Applied Math. & Statistics	Computational Science
Brent Lindquist	SB/Applied Math. & Statistics	Porous Media; Image Analysis
Alfredo Luccio	BNL/AGS	Accelerator Design
Walter Mangel	BNL/Biology	Structural Biology
Glenn Martyna	IBM	Computational Materials Science
Eric Meyer	SB/Physics	Astrophysics
James Muckerman	BNL/Chemistry	Combustion Chemistry
Klaus Mueller	SB/Computer Science	Visualization
John Reinitz	SB/Applied Math. & Statistics	Computational Biology
Robert Rizzo	SB/Applied Math. & Statistics	Computational Biology
Stephen Schwartz	BNL/Environmental Sciences	Aerosol and Climate Modeling
David Sharp	Los Alamos National Lab.	Computational Science
Carlos Simmerling	SB/Chemistry	Protein Structures
William Studier	BNL/Biology	Protein Structures
S. Swaminathan	BNL/Biology	Structural Biology
Douglas Swesty	SB/Physics	Astrophysics
Jin Wang	SB/Chemistry	Computational Biology
Torre Wenaus	BNL/Physics	High Energy and Nuclear Physics
Yan Yu	SB/Applied Math. & Statistics	Uncertainty Quantification
Jincheng Zheng	BNL/Materials Science	Computational Materials Science
Wei Zhu	SB/Applied Math. & Statistics	Statistics; Image Analysis

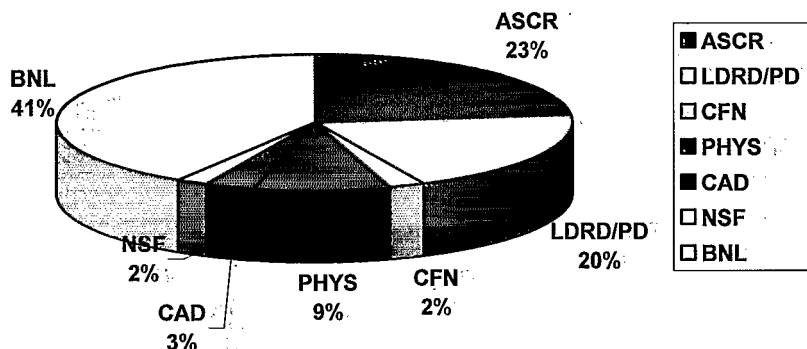
**TABLE 3. STUDENT AFFILIATES**

<b>Name</b>	<b>Affiliation</b>
Joseph Aronson	SB/Instructional Computing
Janet Braunstein	SB/Applied Mathematics & Statistics
Noel Carrascal	SB/Chemistry
Ricky Chachra	SB/Applied Mathematics & Statistics
Xin Chen	SB/Applied Mathematics & Statistics
Yongzhi Chen	SB/Applied Mathematics & Statistics
Yongjun Cheng	SB/Applied Mathematics & Statistics
Roberto Colson	SB/Applied Mathematics & Statistics
Jian Du	SB/Applied Mathematics & Statistics
Alex Dubinsky	SB/Applied Mathematics & Statistics
Srabasti Dutta	SB/Applied Mathematics & Statistics
Bin Fang	SB/Applied Mathematics & Statistics
Yuxiang Gao	SB/Applied Mathematics & Statistics
Youlia Guilman	SB/Physics
Paul Guzzardo	Dowling College
Guowen Han	SB/Applied Mathematics & Statistics
Martin Horstman	SB/Applied Mathematics & Statistics
Zhuying Huang	SB/Applied Mathematics & Statistics
Xiaomei Ji	SB/Applied Mathematics & Statistics
Donghung Kim	SB/Applied Mathematics & Statistics
Noha Lim	SB/Applied Mathematics & Statistics
Xingtao Liu	SB/Applied Mathematics & Statistics
Sudipto Mukherjee	SB/Applied Mathematics & Statistics
Chamainda Kalum Palandage	University of Connecticut
Nitin Pillai	SB/Applied Mathematics & Statistics
Tatiana Polischuk	SB/Applied Mathematics & Statistics
Peter Rissland	SB/ Applied Mathematics & Statistics
Julie Stern	SB/Chemistry
Bentley Strockbine	SB/Pharmacology
Tao Sun	SB/Applied Mathematics & Statistics
Dmitri Volja	SB/Physics
Shuqiang Wang	SB/Applied Mathematics & Statistics
Xuena Wang	SB/Applied Mathematics & Statistics
Perry Woo	SB/Chemistry
Xiangfeng Wu	SB/Applied Mathematics & Statistics
Bin Xu	SB/Applied Mathematics & Statistics
Yue Zhang	SB/Applied Mathematics & Statistics



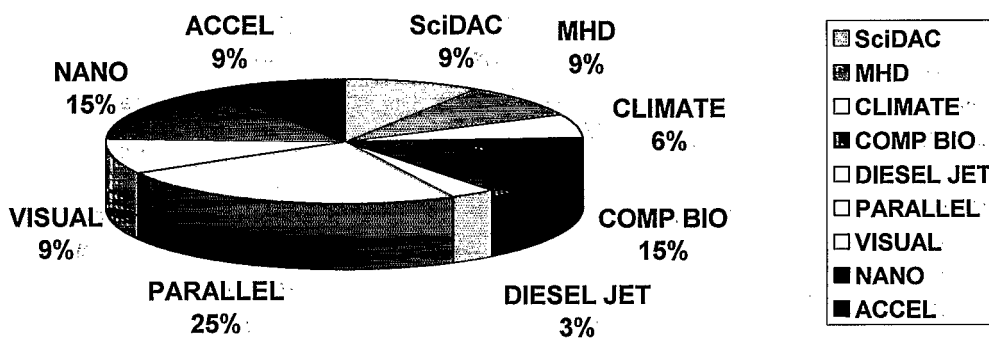
#### 4. RESEARCH

##### CSC Budget FY 2006 Total: \$2,234K



Funding by source: ASCR: DOE Office of Advanced Scientific Computing, LDRD: Laboratory Directed Research and Development, PHYS: BNL Physics, CFN: Center for Functional NanoMaterials, CAD: Collider-Accelerator Department, NSF: National Science Foundation, BNL: Other BNL funding.

##### CSC PROJECTS



## 4.1 SciDAC at CSC

Scientific Discovery through Advanced Computing (SciDAC) was recently renewed as a five-year program to develop the Scientific Computing Software and Hardware Infrastructure to support petascale computations for DOE's research programs.

### 4.1.1 ITAPS

J. Glimm, R. Samulyak, and Z. Xu

At Brookhaven we have formed an alliance with researchers at Argonne, Livermore, Oak Ridge, Pacific Northwest, Rensselaer, Sandia, and Stony Brook, to develop technologies that enable scientists to use complex mesh and discretization strategies easily and interchangeably, within a single simulation on terascale computers. The Interoperable Technologies for Advanced Petascale Simulations (ITAPS) Center, and its predecessor, the Terascale Simulation Tools and Technologies (TSTT) Center, provides interoperable tools to facilitate use of advanced mesh and discretization technologies. The Center has developed standardized interfaces to local mesh refinement codes. Existing codes of the Center partners and new codes are supported to create a plug and play capability, whereby an application user can experiment easily with alternative technologies. Insertion of these tools into targeted applications, including nuclear energy fusion, accelerator design, and ground water modeling, is part of the Center work plan.

CSC work within ITAPS currently focuses on jet breakup, spray formation, bubbly flow for nuclear energy applications, electromagnetic simulations for accelerator design, and on finite element and front-tracking contributions to ITAPS technology.

The ITAPS website is: <http://www.tstt-scidac.org>.

### 4.1.2 TOPS

D. Keyes

Terascale Optimal PDE Simulations (TOPS) is a SciDAC project connected to Brookhaven. TOPS software is being applied in CSC through a collaboration with one of its principal investigators. TOPS deploys a toolkit of open source solvers for partial differential equations, large systems of stiff ordinary differential equations, and linear and nonlinear algebraic systems, including eigenvalue problems, that arise in application areas such as accelerator design, biology, chemistry, magnetohydrodynamics, and particle physics. Scalable solution algorithms -- primarily multilevel methods -- aim to reduce computational bottlenecks by one or more orders of magnitude on terascale computers, enabling scientific simulation on a scale heretofore impossible. Along with usability, robustness, and algorithmic efficiency, an important goal is to attain high computational performance by accommodating to distributed hierarchical memory architectures.

The convergence rates of solvers traditionally employed in PDE-based codes degrade as the size of the system increases. This creates a double jeopardy for applications -- as the cost per iteration grows, so does the number of iterations. Fortunately, the physical structure of PDE problems, such as Poisson's equation for electrostatic potential, provides a natural way to generate a hierarchy of approximate models, through which the required solution may be obtained efficiently.

The efforts defined for TOPS and its collaborations incorporate existing and new optimal algorithms into scientific applications through code interoperability behind a standard interface. TOPS provides support for the software packages Hypr, PETSc (which has powered three

Gordon Bell Prizes in recent years), Sundials, SuperLU, TAO, and Trilinos. Some of these packages are in the hands of thousands of users, who have created a valuable experience base on thousands of different computer systems.

The TOPS project webpage may be found at: <http://www.scidac.gov/math/TOPS.html>

#### **4.1.3 Advanced Computing for 21<sup>st</sup> Century Accelerator Science and Technology** R. Samulyak

The SciDAC Accelerator Modeling Project, "Advanced Computing for 21<sup>st</sup> Century Acceleratory Science and Technology," was initiated in June 2001. Its primary goal is to establish a comprehensive terascale simulation environment for use by the U.S. particle accelerator community. Building upon a previous DOE Grand Challenge project as well as previous individual efforts at several national laboratories and universities, the SciDAC Accelerator Modeling Project represents the largest effort to date for the development of computer codes for accelerator design and analysis. The activities of the project are organized into three application-specific focus areas: Electromagnetics, Beam Dynamics, and Advanced Accelerators. Work in these areas is supported by collaboration with the SciDAC Integrated Software Infrastructure Centers (ISICs) and by personnel (including CSC/BNL) supported through the SciDAC Scientific Application Partnership Program (SAPP).

Research at CSC/BNL is conducted in close collaboration with the Beam Dynamics group. The primary goal of the study is the development of novel mathematical models and software modules for the computation of wake fields and their interaction with particle beams in high intensity accelerators. We have implemented our software in the MaryLie/Integrated Map and Particle Accelerator Tracking code, a parallel code that combines the magnetic optics capabilities based on the Lie algebraic technique with the 3D space charge capabilities and the Synergia framework.

The SciDAC website is <http://www.osti.gov/scidac>.

#### **4.1.4 Statistical Approaches to Aerosol Dynamics for Climate Simulation** R. McGraw and W. Zhu

The SciDAC climate modeling project, "Statistical Approaches to Aerosol Dynamics for Climate Simulation," was recently awarded five years funding.

This research supports the goal of the DOE Climate Change Prediction Program (CCPP), which is to determine the range of possible climate changes over the 21<sup>st</sup> century and beyond using a more accurate climate system model. In partnership with the Applied Mathematics and Statistics Department at SUNY Stony Brook, we develop new statistical approaches for improving the representation of aerosols, aerosol microphysical processes, and aerosol-cloud interactions in the Community Climate System Model (CCSM). The new approaches, which track multivariate moments of the particle population, are highly efficient, yet provide the comprehensive representation of natural and anthropogenic aerosols, and of their mixing states and direct and indirect effects, that the CCSM will require. Findings from this study will be incorporated into a new aerosol microphysical module in time for CCSM5.

## 4.2 BIOLOGICAL AND ENVIRONMENTAL RESEARCH

Novel algorithms and hardware allow fine-scaled parallelism for all atom simulations to model structural biology. We are emphasizing conformational changes in proteins with known structures, in collaboration with BNL biologists.

Brookhaven has strong experimental programs in atmospheric chemistry and brain physiology. Both programs are data intensive. The challenge is not only data quantity, but especially data understanding. We have developed new tools for interactive visual and statistical data analysis. We are actively using these tools in diverse applications. See also Sections 4.3.12 and 4.3.13.

### 4.2.1 Binding of Anti-Fusion Peptides with HIVgp-41

B. Strockbine, S. Mukherjee, N. Carrascal, and R.C. Rizzo

Infection of a target cell by a virus requires fusion of the viral envelope and the target cell plasma membrane. Fusion events during infection by human immunodeficiency virus (HIV) are mediated by a viral glycoprotein called HIVgp41. Prior experimental studies of the core domain of HIVgp41 have revealed a structure containing a trimer-of-hairpin motif in which three outer C-helices loop and wrap around three inner N-helices [1]. During fusion, a proposed intermediate state transiently exposes both inner and outer helices. Inhibitors designed to bind to the inner N-helices can prevent reformation of the fusogenic structure thereby inhibiting virus-host cell membrane fusion [2,3]. In 2003 the FDA approved the first inhibitor in this anti-fusion class, Fuzeon (T20), a 36 amino acid C-helix peptide mimic which specifically targets gp41 [4].

In this work we use computational methods to model binding of "second generation" anti-fusion peptides complexed with HIVgp41 and use the accompanying energetic and structural results to estimate binding affinities for comparison with reported experimental activities [5]. These peptides are of a different sequence than T20 and contain functionality which overlaps with a highly conserved hydrophobic pocket described by Kim and coworkers [1]. The pocket could be exploited both in the design of improved anti-fusion peptides and for development of low molecular weight inhibitors. Figure 1 shows how two Trp residues and one Ile on second generation C-peptides pack into the interface formed by the inner N-helices [6,7]. These conserved pocket interactions are thought to play important roles in stabilizing the fusogenic conformation of HIVgp41. The ability to delineate which interactions are most important will ultimately enable the design of better inhibitors with improved ability to combat resistance mutations.

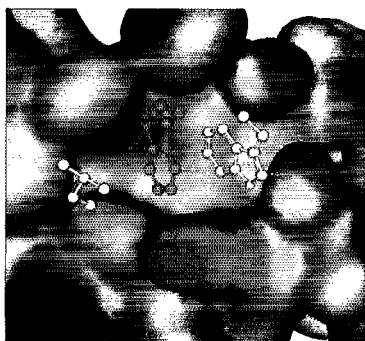


Figure 1. A surface model of the inner helices of HIVgp41 showing the hydrophobic pocket with the partnering residues from the C-helix shown as ball and stick models.

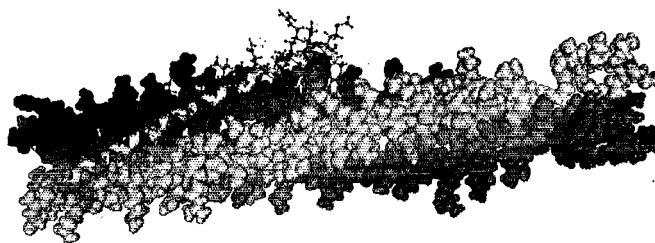


Figure 2. HIVgp41 inner N-helices represented in blue all atom models with the C34 C-helix peptide represented in red ball and stick with the parent-Trp residue in the hydrophobic shown in green.

Chan et al. [6] have shown that the ability to inhibit fusion directly correlates with melting temperatures ( $T_m$ ) for complexes containing C-peptides that interact directly with the hydrophobic pocket. Experiments with a peptide termed C34 (Figure 2 red peptide) revealed that changing the parent Trp (green residue Figure 1) at position 631 to successively smaller residues reduced the ability of the analogs to inhibit both viral entry and cell membrane fusion [6]. A strong linear correlation between  $T_m$  of complexes and the log of  $IC_{50}$  values was observed, which provides compelling evidence that the experimental activities are a measure of peptide affinity for the receptor, and that binding of C34 peptides containing the pocket region is responsible for the observed inhibition [6]. In this work we use all-atom molecular dynamics (MD) simulations followed by MM-GBSA (Molecular Mechanics Generalized Born Surface Area) energy analyses to probe structure-activity relationships (SAR) for a series of six C-peptides and test current hypotheses about the importance of pocket residues for modulating binding. A well tested computational model for this system will assist the development of improved second generation anti-fusion peptides, and will be useful for *in silico* high-throughput screening (docking) of potential small molecule inhibitors.

Docking will be used to predict how compounds interact with the gp41 pocket by considering thousands of different conformations, orientations, and the electrostatic and steric complementarity of each protein-ligand complex. As contributors to the DOCK program [8], we have recently optimized procedures [9] for computing important desolvation terms associated with molecular recognition. Specifically, we have added new algorithms to DOCK, based on MM-GBSA methods, to more accurately estimate free energies of binding [10,11]. Experimental testing for compounds suggested to bind tightly to gp41 will be done in collaboration with Dr. Miriam Gochin at the University of California at San Francisco (San Francisco, California). Compounds suggested by computational docking and determined experimentally to bind will be analyzed with *in vitro* syncytium and HIV viral infectivity assays. The long-term goal of Dr. Rizzo's research is to develop promising non-peptide drug candidates for HIV-1 fusion inhibition. We gratefully acknowledge support from the Department of Applied Mathematics at Stony Brook, the NYSTAR James D. Watson Investigator Program, and the Computational Science Center at Brookhaven National Laboratory.

## References

- [1] Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89** (2): 263-273 (1997).
- [2] Eckert, D.M. and Kim, P.S. Mechanisms of viral membrane fusion and its inhibition. *Annu. Rev. Biochem.* **70**: 777-810 (2001).
- [3] Jiang, S., Lin, K., Strick, N., and Neurath, A.R., HIV-1 inhibition by a peptide. *Nature* **365** (6442): 113 (1993).
- [4] Poveda, E., Briz, V., and Soriano, V. Enfuvirtide, the first fusion inhibitor to treat HIV infection. *Aids Reviews* **7** (3): 139-147 (2005).
- [5] Strockbine, B. and Rizzo, R.C. Binding of anti-fusion peptides with HIVgp41 from molecular dynamics simulations: Quantitative correlation with experiment. *Proteins: Struct., Funct., Bioinf.* Submitted, 2006.
- [6] Chan, D.C., Chutkowski, C.T., and Kim, P.S. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc. Natl. Acad. Sci. U.S.A.* **95** (26): 15613-15617 (1998).

- [7] Jiang, S. and Debnath, A.K. A salt bridge between an N-terminal coiled coil of gp41 and an antiviral agent targeted to the gp41 core is important for anti-HIV-1 activity. *Biochem. Biophys. Res. Commun.* **270** (1): 153-157 (2000).
- [8] Moustakas, D.T., Therese Lang, P.T., Pegg, S., Pettersen, E., Kuntz, I.D., Broojimans, N., and Rizzo, R.C. Development and validation of a modular, extensible docking program: DOCK-5. *J. Comput. Aided Mol. Des.* Submitted, 2006.
- [9] Rizzo, R.C., Aynechi, T., Case, D.A., and Kuntz, I.D. Estimation of absolute free energies of hydration using continuum methods: Accuracy of partial charge models and optimization of nonpolar contributions. *J. Chem. Theory Comp.* In press, 2005.
- [10] Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A., and Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Amer. Chem. Soc.* **120** (37): 9401-9409 (1998).
- [11] Massova, I. and Kollman, P.A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design* **18**: 113-135 (2000).

#### 4.2.2 Molecular Dynamics Study of HIV-Protease

V. Hornak, A. Okur, R.C. Rizzo, and C. Simmerling

Due to its central role in processing viral polypeptide precursors, HIV-1 protease (HIV-PR) continues to be one of the primary targets of anti-AIDS drug discovery. The introduction of HIV-1 protease (HIV-PR) inhibitors has led to a dramatic increase in patient survival; however, these gains are threatened by the emergence of multi-drug resistant strains. Design of inhibitors that overcome resistance would be greatly facilitated by deeper insight into the mechanistic events associated with binding of substrates and inhibitors, as well as an understanding of the effects of resistance mutations on the structure and dynamic behavior of HIV-PR. An extensive set of X-ray crystal structures of HIV-1 protease has been solved [1], revealing a C2 symmetric homodimer with a large substrate binding pocket covered by two glycine rich  $\beta$ -hairpins, or flaps [2-4]. Consistent structural differences are present between the bound and free states of the protein. In all of the liganded forms, the flaps are pulled in towards the bottom of the active site ("closed" form), while the structures for the unbound enzyme all adopt a "semi-open" conformation with the flaps shifted away from the catalytic site, but still substantially closed over the active site and in contact with each other. Although large-scale flap opening is presumably required for normal substrate access to the active site, no crystallographic structures representing such an open configuration have been reported and thus the mechanism of inhibitor entry and binding remains unknown.

In contrast to the static view provided by crystallography, molecular dynamics (MD) simulations can provide valuable insight into time-dependent structural variation. We performed unrestrained, all-atom molecular dynamics simulations of HIV-PR that sampled large conformational changes of the active site flaps [5]. The unliganded protease underwent spontaneous and reproducible conversions between the "closed" and "semi-open" forms observed in crystal structures of inhibitor-bound and unliganded protease, respectively. Simulations in the presence of a cyclic urea inhibitor yield stable closed flaps. Furthermore, we observed several events in which the flaps of the unliganded protease opened to a much greater degree than observed in crystal structures and subsequently returned to the semi-open state. Our data strongly support the hypothesis that the unliganded protease predominantly populates the semi-open conformation, with closed and fully

open structures being a minor component of the overall ensemble. The results also provide a model for the flap opening and closing that is considered to be essential to enzyme function.



Figure 3. Time sequence showing the opening and closing of the flap in HIV-1 protease.

Further computational studies investigated whether the model for the open state based on our simulations was capable of binding to inhibitors and adopting the closed form. We manually placed an inhibitor in the proximity of the flaps (see Figure 3). During subsequent simulations, the inhibitor reproducibly induced spontaneous conversion to the closed form as seen in all inhibitor-bound HIV-PR crystal structures, with root mean square deviation (RMSD) of  $\sim 1$  Å from the crystal structure of the complex for the inhibitor and each flap despite initial RMSD values of 6 - 11 Å [6]. The results demonstrate that all-atom simulations have the ability to significantly improve poorly docked ligand conformations and reproduce large-scale receptor conformational changes that occur upon binding.

## References

- [1] Vondrasek, J. and Wlodawer, A. *Proteins-Structure Function and Genetics* **49**: 429-431 (2002).
- [2] Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., and Springer, J.P. *Nature* **337**: 615-620 (1989).
- [3] Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., and Kent, S.B.H. *Science* **245**: 616-621 (1989).
- [4] Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E., Geoghegan, K.F., Hawrylik, S.J., Lee, S.E., Scheld, K.G., and Hobart, P.M. *Nature* **342**: 299-302 (1989).
- [5] Hornak, V., Okur, A., Rizzo, R. and Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Nat. Acad. Sci. USA* **103**: 95-920 (2006).
- [6] Hornak, V., Okur, A., Rizzo, R. and Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.* **128**: 2812 (2006).

### 4.2.3 MD Simulations of the Activation of the Adenovirus Proteinase

J.V. Stern, L. Slate, J.W. Davenport, W.J. McGrath, and W.F. Mangel

The adenovirus proteinase (AVP) is a relatively inactive enzyme that can be activated by an 11-amino-acid peptide cofactor, pVIc [1,2]. The binding of pVIc increases the activity of AVP more than 3500-fold. The crystal structure of the AVP-pVIc complex has been determined experimentally using x-ray diffraction and the pVIc peptide was found to bind to AVP in an

extended conformation quite far from the active site, as shown in Figure 4 [3]. The peptide traverses the surface of the protein, ranging from 14 Å to more than 30 Å away from the active site. This raises the question: how can the peptide at such a large distance from the active site have such a profound influence in the functioning of the enzyme?



Figure 4: Crystal structure of AVP-pVlc (1NLN). The active site His54 and Cys122 are in green. The 11 amino acid peptide, pVlc is the yellow band at the bottom. The waters from the crystal structure are shown in orange.

Recently, the crystal structure of AVP in the absence of the cofactor was solved as well. Comparison of this structure to the structure of the AVP-pVlc complex is revealing how pVlc activates the enzyme. We hypothesize that in order for the binding of pVlc to activate the enzyme, a signal must be transduced by a sequence of structural changes beginning at the pVlc binding site on AVP and ending in the active site of AVP. Using molecular dynamics simulations, it is hoped that the sequence of structural changes that occur can be understood and exploited to provide targets for drugs. If the binding of a drug to a target within the sequence of structural changes can prevent the signal from being transduced, therefore preventing activation, the drug could be an effective anti-viral agent.

We have performed several sets of molecular dynamics simulations on this protein. In the first, using a generalized Born implicit water model, the peptide was removed from the known structure of the AVP-pVlc complex and the structure was equilibrated at 310K using a series of molecular dynamics and energy minimization steps. Following that, an 11 nanosecond simulation led to a new structure in which the geometry of the 4 residue active site was significantly expanded. In fact the distance from HIS 52 to CYS 120 increased by 1.4 Å. We hypothesize that this increase is the immediate cause of the deactivation of the protease.

In the second set, an explicit water model was used [4]. Simulations were started from the experimental AVP structure (without the pVlc) to which a pVlc peptide had been docked. These simulations were run for ~3.7 nsec. They illustrate further changes in the backbone structure including a most interesting helix-coil transition. Using this approach, one possible step in the activation mechanism has been observed. A third set of simulations was run starting with the known AVP-pVlc complex but with the disulfide bond that binds pVlc to AVP removed. The objective was to test a model describing how pVlc binds to AVP by inducing a pocket in AVP, another possible step in the activation mechanism. A simulation proceeding in the reverse direction, unbinding, should therefore show some loss of the induced pocket. After ~100 picoseconds the last three amino acids of PVlc were indeed seen to leave the induced pocket.



The explicit water model molecular dynamics (MD) simulations were performed using NAMD [5], a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. AMBER force fields [6] were used throughout. Our protein contains 204 amino acids plus the 11 amino acid peptide cofactor. After protonation this constitutes about 3500 atoms. Adding water yields a total of 30,000 atoms. The implicit water model MD simulations used a parameterization of the "OBC" model of generalized Born [7], the AMBER force fields, and the AMBER parallel molecular dynamics code [8]. Langevin dynamics with a collision frequency of one inverse picosecond was employed to maintain a constant temperature of 310 K throughout the MD run. The fastest motion in the system, i.e. the bond stretching freedom, was removed for bonds involving Hydrogen using the SHAKE algorithm, thereby allowing a larger timestep (1 femtosecond) to be used.

We plan to continue these simulations to elucidate the transformations involved in activating the enzyme and to discover sites along the pathway that may serve as targets for drugs that will act as antiviral agents.

## References

- [1] Mangel, W.F., McGrath, W.J. et al. Viral DNA and a viral peptide can act as cofactors of adenovirus virion proteinase activity. *Nature* **361**: 274-275 (1993).
- [2] Mangel, W.F., Baniecki, M.L., and McGrath, W.J. Specific interactions of the adenovirus proteinase with the viral DNA, an 11-amino-acid viral peptide, and the cellular protein actin. *Cell. Mol. Life Sci.* **60**: 2347-2355 (2003).
- [3] Ding, J., McGrath, W.J., et al. Crystal structure of the human adenovirus proteinase with its 11 amino-acid cofactor. *EMBO J.* **15**: 1778-1783 (1996).
- [4] Jorgensen, W.L., Chandrasekhar, J. et al. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**: 926-935 (1983).
- [5] Laxmikant, K.R., et al. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Physics* **151**: 283-312 (1999).
- [6] Cornell, W.D., Cieplak, P. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**: 5179-5197 (1995).
- [7] Onufriev, A., Bashford, D., and Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* (2004).
- [8] Pearlman, D.A., Case, D.A. et al. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Physics Commun.* **91**: 1-41 (1995).

### 4.2.4 MD Simulations of Biotoxins and of PDZ Domains

X. Chen and Y. Deng

The toxin produced by the bacterium *Clostridium botulinum* is one of the deadliest known [1]. Recently, a great deal of information about its mode of action has become available through structural and other biophysical studies [2]. The toxin itself (BoNT) is a protein of approximately 1300 residues, with two chains, a light chain (LC) of approximately 50 kDa and a heavy chain (HC)

of approximately 100 kDa. The two are linked by a disulfide bond. The LC contains a  $Zn^{2+}$  ion, which acts within the cell to cleave a protein necessary for neurotransmission leading to paralysis and death. This process is believed to involve several steps: binding of the toxin to the endosomal membrane; translocation through the lipid bilayer; and proteolysis of specific neurotransmitters within the cell. Structural studies show that the HC contains two domains responsible for binding and translocation, while the LC contains the catalytic domain. Both the structure of the toxin outside the cell (but not at endosomal pH) and the mechanism of action of the zinc protease are reasonably well understood. But, the structures at low pH and the mechanism of the translocation remain illusive.

The results of 0.2 micro-seconds simulated time for two different temperatures and pH values are shown in Figure 5:

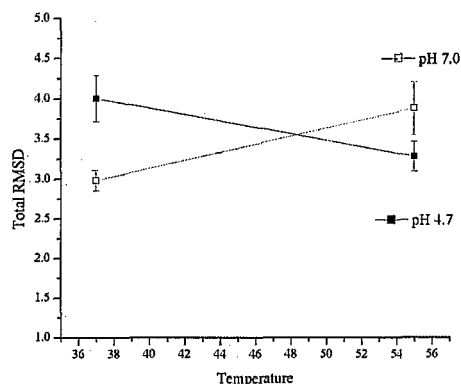


Figure 5. Comparison of the RMSD for the whole protein.

We also investigated the coupling between the histidine residues with the zinc ion as in Figure 6.

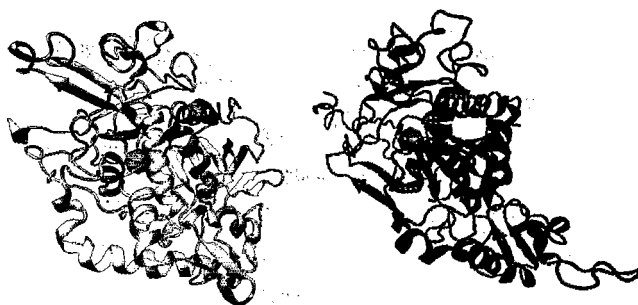


Figure 6. Structures of LC in LC only run at different pH and 37 °C, Red: pH 4.7; Blue: pH 7.0.

It is well known that the pH dependence of these structures is not fully described by the traditional method [3]. The protonation state is a function of pH, but the change with pH depends on the protonation state, requiring a self-consistent solution. A recent new approach combining molecular dynamics and Monte Carlo simulations has been developed [4]. We have implemented this method with the AMBER 8 [5] package to verify our results. [6]

We also simulated the disheveled PDZ domain with ligands [7]. The Disheveled (Dvl) PDZ domain is believed to play an essential role in the canonical and noncanonical Wnt signaling pathways, which are involved in embryo development as well as in tumorigenesis. Also, it binds directly to Frizzled (Fz) receptors. An organic molecule (NSC668036) from the National Cancer Institute small-molecule library has been identified to be able to bind to the Dvl PDZ domain [8]. We

developed molecular dynamic simulations to analyze the binding between them and analyzed their interactions in details.

Interesting phenomena were revealed by our simulations. As in Figure 7, we found the convergence of the 30 structures of the NSC668036. The figure uses colors to represent the relative RMSD of different structures after 10ns simulation. Since the structures have zero RMSD when comparing with itself, the diagonal line of the matrix is zero. An island with small numbers in the matrix means that the relevant structures have similar conformation, and thus, tend to converge to a similar structure.

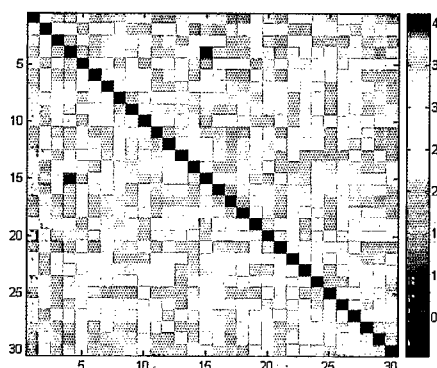


Figure 7. Final state of PDZ domain after 10 ns simulation.

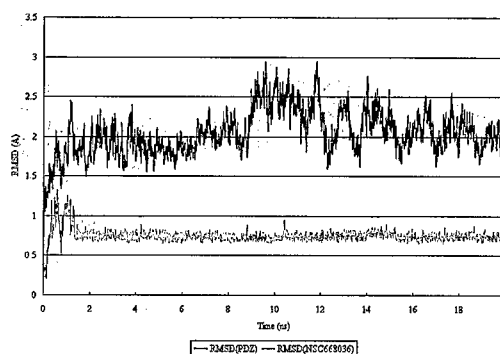


Figure 8. RMSD of PDZ domain and ligand #10.

We also found some of the 30 structures are more stable after docking with the PDZ domain. For example, if we take a close look at structure #10, which is shown in Figure 8, we can find that the ligand remains quite stable after the docking, which is indicated by its low and stable RMSD.

## References

- [1] Singh, B.R. *Nature Structural Biology* **7**: 617-619 (2000).
- [2] Chen, X. and Deng, Y. *J. Molecular Modeling*. Conditionally accepted, 2006.
- [3] Smith, L.J. et al. *Proteins* **36**: 77 (1999).
- [4] Mongan, J. et al. *J. Comp. Chem.* **25**: 2038 (2004).
- [5] Pearlman, D.A. et al. *Comp. Phys. Comm.* **91**: 1-41 (1995).
- [6] Chen, Y., Chen, X., and Deng, Y. *J. Comp. Phys.* To be submitted, 2006.
- [7] Chen, X. and Deng, Y. *J. Molecular Modeling*. To be submitted, 2006.
- [8] Shan, J. et al. *Biochemistry* **44(47)**: 15495-15503 (2005).

## 4.2.5 Visualization and Data Mining for Detection of Cancer

W. Zhu, R. Bennett, J. Kovach, and M. McGuigan

The statistical problem of cancer detection is addressed using diagnostic data, that is, blood samples from patients with known diagnosis, having ovarian cancer or normal [1,2]. When statistical tests are applied to the ovarian cancer data, accuracy in the range of 80-95% results. More recently, we have applied the same technology for the detection of head and neck cancer. Again, accuracy in the range of 80-95% results [3].

We are developing an open source toolkit for NIH/FDA/NCI to assist with the processing and analysis of datasets derived from proteomic experiments [4,5,6]. Open source software provides a mechanism for leveraging existing toolkits, sharing expertise and accelerating development. Our toolkit utilizes components from several existing open source projects. These include the visualization toolkit (VTK) and the machine learning in C++ library (MLC++). It will initially be customized for serum proteomic pattern diagnostics involving the monitoring for ovarian cancer recurrence.

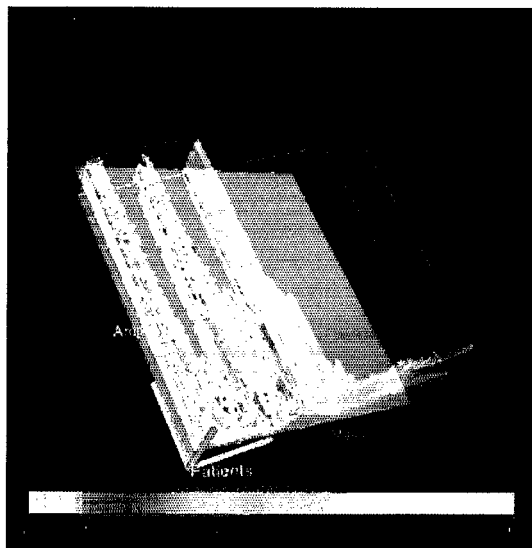


Figure 9. Splat visualization displaying serum proteomic patterns from 22 patients with ovarian cancer.

Figure 9 shows a Splat Visualization rendering of serum proteomic data from 22 patients following processing by a high-resolution mass spectrometer. The majority of the spectra are from patients with known ovarian cancer. The Splat Visualizer tool is an aggregate 3-D plotter that allows rapid manipulation and viewing of large datasets. The tilted x-axis (labeled *Mass*) represents the mass values of the peptides and protein fragments. The y-axis (labeled *Amp*) shows normalized amplitude values (relative measure of abundance). The z-axis (labeled *Patients*) has been used to stack individual patient spectra, thus allowing comparison. The graphic has also been automatically colored as a function of support (number of data points behind the voxel), white and blue having the least amount of support, and red the most. Therefore, the patient's unique proteomic signature can be compared and contrasted with known controls and calibrators. Figure 10 displays output of the Multi-view visualization tool. This tool can be thought of as a superset of the Splat Visualizer, since it allows for either synchronous or asynchronous coordination of multiple windows, all containing different proteomic datasets. In all four views the x- and y-axes are defined as in Figure 10. However, each view employs a different clinical parameter in the z-axis, allowing simultaneous proteomic pattern viewing, but in that window's particular context.

In each window, serum proteomic spectra from approximately 100 patients with known ovarian cancer have been aggregated and grouped according to a clinical parameter of interest. Beginning top left and moving clockwise: spectra are rendered as a function of the patient's age, next spectra are grouped and displayed according to tumor grade, next by a lab value [Cancer Antigen 125 (CA 125)], and finally by stage of disease. Any window can be maximized and subsequent operations performed, such as: zoom, pan, rotate, drill, or execution of a defined command.

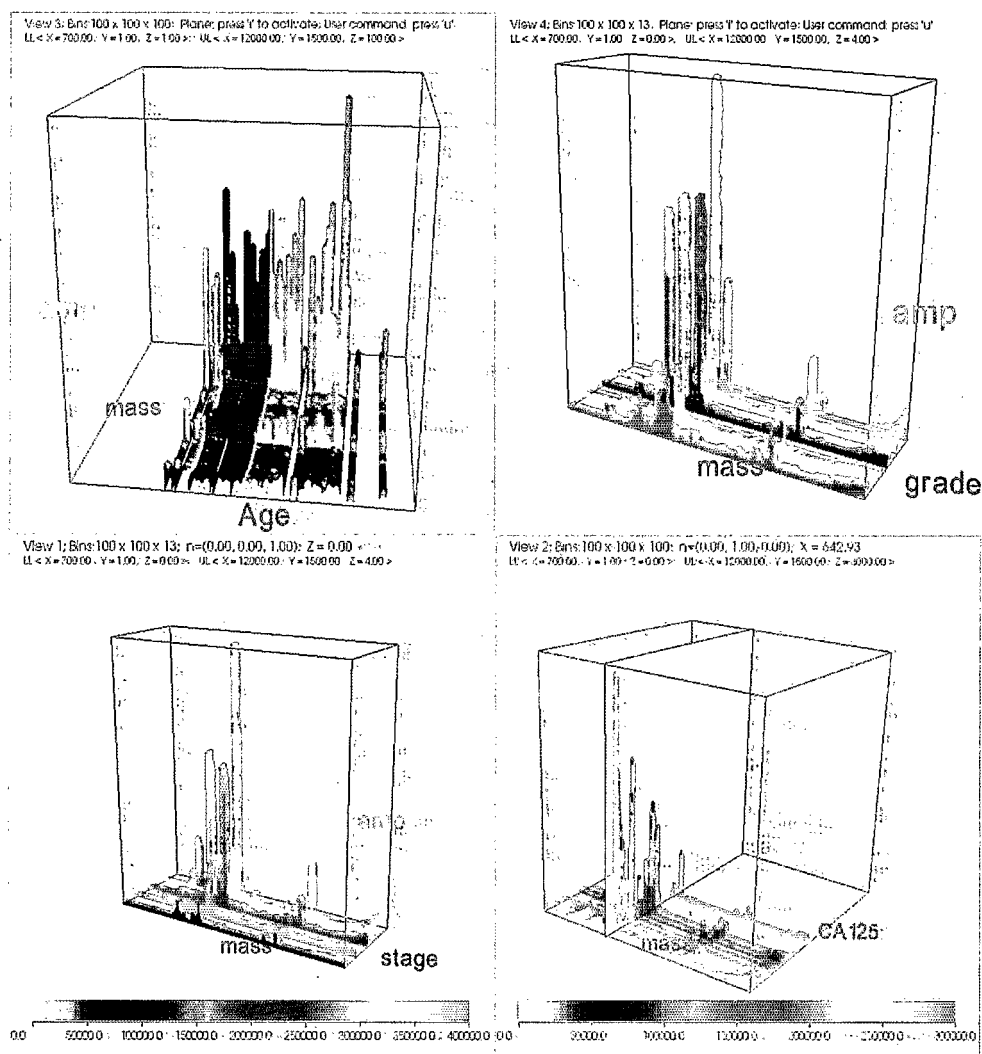


Figure10: Multi-view visualization of aggregate proteomic patterns (approximately 100 patients, most with ovarian cancer) viewed as a function of a clinical parameter.

## References

- [1] Petricoin III, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Rusaro, V.A., Steinberg, S.M, Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., and Liotta, L.A. *Lancet* **359**: 572-577 (2000) and *Lancet* **360**: 169-171 (2002).
- [2] Zhu, W., Wang, X., Ma, Y., Rao, M., Glimm, J., and Kovach, J. Detection of cancer specific markers amidst massive mass spectral data. *Proc. Nat. Acad. Sci.* **100**: 14666-14671 (2003).
- [3] Wang, X., Zhu, W., Pradhan, P., Ji, C., Ma, Y., Semmes, J., Glimm, J., and Mitchell, J. Feature extraction in the analysis of proteomic mass spectra. *Proteomics* **6**: 2095-2100 (2006).
- [4] Johann, D.J., McGuigan, M.D., Tomov, S., Fusaro, V.A., Ross, S., Conrads, T.P., Veenstra, T.D., Fishman, D.A., Whiteley, G.R., Petricoin, E.F., and Liotta, L.A. Novel approaches to visualization and data mining reveal diagnostic information in the low amplitude region of serum mass spectra from ovarian cancer patients. *Disease Markers* **19**: 197-207 (2004).

- [5] Johann, D.J., McGuigan, M.D., Tomov, S., Blum, E., Whiteley, G.R., Petricoin, E.F., and Liotta, L.A. Toward a Systems Biology Software Toolkit, 17th IEEE Symposium on Computer-Based Medical Systems, June 2004.
- [6] Johann, D.J., McGuigan, M.D., Patel, A.R., Tomov, S., Ross, S., Conrads, T.P., Veenstra, T.D., Fishman, D.A., Whiteley, G.R., Petricoin, E.F., and Liotta, L.A. Clinical proteomics and biomarker discovery. *Annals of the New York Academy of Sciences* **1022**: 295-306 (2004).
- [7] Wang, X., Zhu, W., Pradhan, K., Ji, C., Ma, Y., Semmes, O.J., Glimm, J., and Mitchell, J. Feature extraction in the analysis of proteomic mass spectra. *Proteomics* **6(7)**: 2095-2100 (2006).

#### 4.2.6 Atmospheric Aerosols

W. Zhu, K. Mueller, and R. McGraw

The purpose of this project is to develop data processing tools for the classification and time series analysis of atmospheric aerosol particles. Atmospheric aerosols play an important role in both climate and public health. Current research is focused on determining the chemical composition of aerosols, and thus the composition of air and its evolution over time. Massive mass spectra data, on the scale of terabytes per week, have been collected continuously over time by researchers in the Atmospheric Sciences Division at Brookhaven, using their cutting-edge, field deployable, high-precision mass spectrometer. The first task is to determine the chemical composition of each aerosol based on its mass spectrum. To achieve this, we classify aerosols by their chemical compositions and thus determine the air composition at each time point. Finally, we study the time series evolution of atmospheric compositions identifying normal as well as abnormal patterns in real time.

To convey the analysis results to the scientists in an intuitive way and to incorporate their expert knowledge into the data analysis phase, an interactive graphical user interface utilizing modern scientific visualization techniques becomes a necessity. Large-scale BNL data collections have reached sizes for which straightforward visualization techniques are starting to fail. In this spirit, our proposed approach couples a powerful Bayesian classification and multivariate time series analysis engine with an intuitive and responsive graphical user interface to fine-tune the underlying models. We call this application *SpectrumMiner*, a significant component of a more general visual data mining framework and application, called ViStA, which we are currently developing at BNL [1,2,3,4] (for more detail see Section 4.3.12).

**The statistics engine.** We first classify aerosols based on their mass spectra through an iteration of expert-machine interaction using a Bayesian classifier. The Bayesian classifier can incorporate not only explicit classification rules but also prior knowledge in terms of a partial training data set. We have also been implementing automatic procedures to elucidate the structure of the compounds. To study the evolution of aerosols and the changes in atmospheric composition versus time, both the univariate and the multivariate time series analyses are employed to unravel trends and patterns.

The traditional statistical classifier would take only explicit classification rules. Machine learning techniques such as the neural network would take only a training set. However, the semi-supervised clustering approach we are taking requires the classifier to learn from both the explicit rules and the implicit rules embedded in a partial training set established by the experts as they survey the current clustering results and make adjustments to the clusters. For this we have implemented an interactive Bayesian classification framework that could absorb the up-to-date prior information in both the explicit and implicit formats, and produce the updated posterior classification results.

For efficient classification and automatic structure elucidation, we have been constructing a molecule library where the signature profile and class membership of each molecule is established. The molecules are classified along a natural chemical classification tree with two categories--organic and inorganic--at the initial node. Subsequently, the organics are further divided into classes of carboxylic acids, aldehydes, ketones, alkenes, alkanes, aromatics, etc. We begin with the spectra of known molecules (NIST library or lab-generated).

**Visualization.** The statistical analysis engine is combined with a highly visual interface to facilitate interactive exploration, mining, classification and survey of these large, high-dimensional data collections. In order to empower scientists to control and fine-tune the mining and classification process in an intuitive and interactive way, SpectrumMiner's hierarchical classification algorithm is user-steerable via a novel multi-modal visual interface. An important component of this interface is the *interactive dendrogram*, where hierarchy nodes are placed on concentric circles whose radii are determined by the dissimilarity of the node's sub-tree. We chose a circular layout of the dendrogram since it makes better use of space than its linear counterpart. It inherently dedicates less drawing space to the higher-level, less numerous nodes, and distributes more space to the many leaf nodes along the circumference of the circle. See Figure 11, where we show a screen capture of SpectrumMiner, with the interactive dendrogram located on the bottom right. Edges are colored using a rainbow colormap to indicate the number of data items they carry. To the left of the dendrogram is the node viewer. Selecting a particular node will display the average spectrum of all data items classified into the node (window with white background), as well as the node's data composition (window with blue background just below). We are currently incorporating the classification steering capabilities into our system. By inspecting the present classification in the node viewer and the dendrogram, scientists may decide that some particles, or the entire node, have been misclassified. To correct this error, they then simply drag the concerned particles or node into the proper location within the hierarchy, which subsequently triggers a refinement of the classification rules.

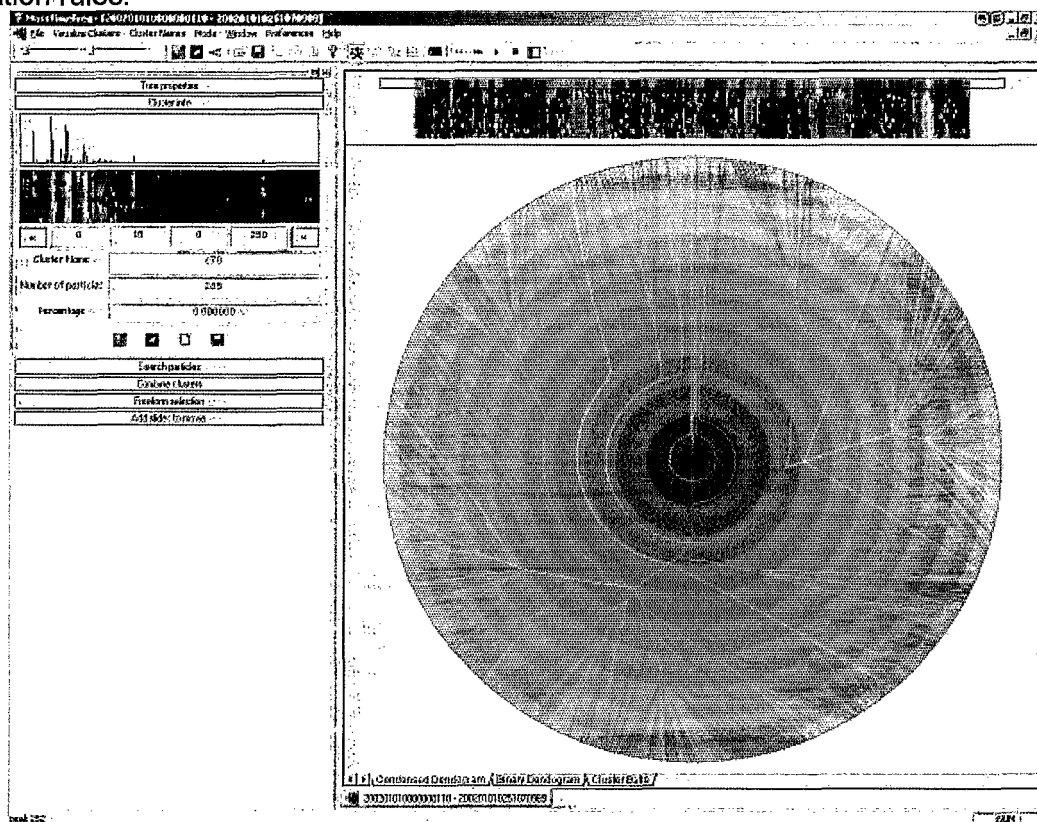


Figure 11: SpectrumMiner's interactive dendrogram interface.

## References

- [1] Imrich, P., Mueller, K., Imre, D., Zelenyuk, A., and Zhu, W. A Hardware- Accelerated Rubbersheet Focus + Context Technique for Radial Dendrograms. IEEE Information Visualization Symposium'03; Seattle, October 2003.
- [2] Imrich, P., Mueller, K., Imre, D., Zelenyuk, A., and Zhu, W. 3D ThemeRiver. IEEE Information Visualization Symposium'03; Seattle, October 2003.
- [3] Yoon, C. and McGraw, R. Representation of generally mixed multivariate aerosols by the quadrature method of moments: I. Statistical foundation. J. Aerosol Sci. **35**: 561-576 (2004).
- [4] Yoon, C. and McGraw, R. Representation of generally mixed multivariate aerosols by the quadrature method of moments: II. Statistical foundation. J. Aerosol Sci. **35**: 577-598 (2004).

## 4.3 ADVANCED SCIENTIFIC COMPUTING

### 4.3.1 ITAPS: Atomization and Spray

Z. Xu, J. Glimm, X. Li, R. Samulyak, and C. Tzanos

In the past year, we have studied mechanisms leading to jet breakup and spray formation for a high-speed diesel jet injected through a circular nozzle using first principles simulations. The simulations are conducted using the Front-Tracking method within a 2D axis-symmetric geometry. Our goal is to model the spray at a microphysical level, with the creation of individual droplets. Through our study, we have found that the formation of cavitation vapor bubbles is the key phenomenon contributing to the breakup. During the inception stage of the cavitation, the flow is characterized by traveling bubbles. The bubbles grow intensively and coalesce into big bubbles. These bubbles create an atomizing spray.

We have developed a heterogeneous EOS model consisting of explicitly tracked cavitation vapor bubbles within the liquid diesel to describe the mixed phase (diesel vapor/liquid) regime. Motivated by homogeneous nucleation theory, we have formulated a dynamic bubble insertion algorithm to simulate the incipient creation of cavitation bubbles.

The interface that separates the liquid and the vapor is modeled as a phase boundary. A new phase transition model to calculate the solutions for the dynamic phase transitions has also been developed. In this model, viscosity and the surface tension on the phase boundary are neglected, because the thermal effects are normally dominant over the viscous effects in phase transitions. We use the kinetic theory of evaporation to give the evaporation rate with a coefficient  $\alpha$  determined experimentally. The net mass flux of evaporation then is  $M_{ev} = \alpha (P_{sat}(T) - p_v) / (\sqrt{2\pi RT})$ . On the phase boundary, the temperature is continuous, while the vapor pressure is allowed to have a deviation from the Clausius-Clapeyron equation. Thus the interface motion depends on the phase change under nonequilibrium thermodynamic conditions along with hydrodynamic conditions.

Three active scales enter into this simulation. The smallest (nm) is a thermal layer at the edge of the vapor bubble to mediate the phase transition rate. This scale is modeled, not resolved. The next two scales are the bubble diameter (mm) and the nozzle/combustion chamber (cm), both resolved numerically.



In the simulations, we used n-heptane as a replacement for No. 2 diesel fuel because the thermal data of No. 2 diesel fuel is not available, and N-heptane is the major component of No. 2 diesel fuel. We have compared our simulation results with the experimental data [1-4]. These experimental data present mass vs. time in a 0.55 mm wide observation window that is centered 1 mm from the nozzle exit. Like the experiments, the simulations also predicted a peak mass. After the peak, the predicted mass exhibits a small variability (Figure 12). The opening angle of the jet (Figure 13), which varies as a function of time, is about 15 to 30 degrees and is in agreement with its experimental value. Figure 14 shows a comparison of jet tip velocity computed from simulations [5,6] with experimental data. Although the value computed from simulation predictions exhibits a wide variability, on the average, it is in agreement with the value computed from the experiment. It should be noted that the experimental data has been averaged over 100 injection cycles to remove fluctuations.

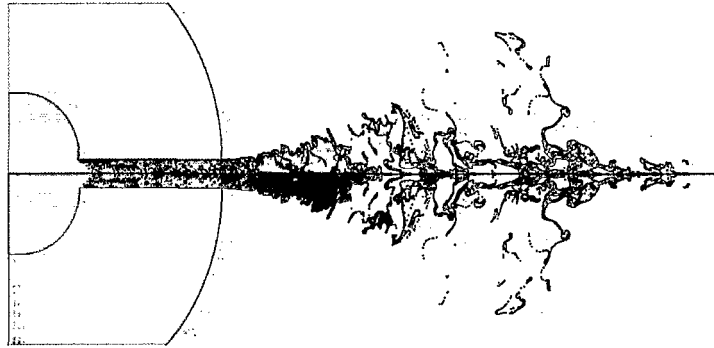


Figure 12. The vorticity (above) vs. density (below). The dark spots are the diesel vapor bubbles. The bubbles appear and disappear as rarefaction waves progress up and down the nozzle. The large blue region in the lower frame is a vapor region.

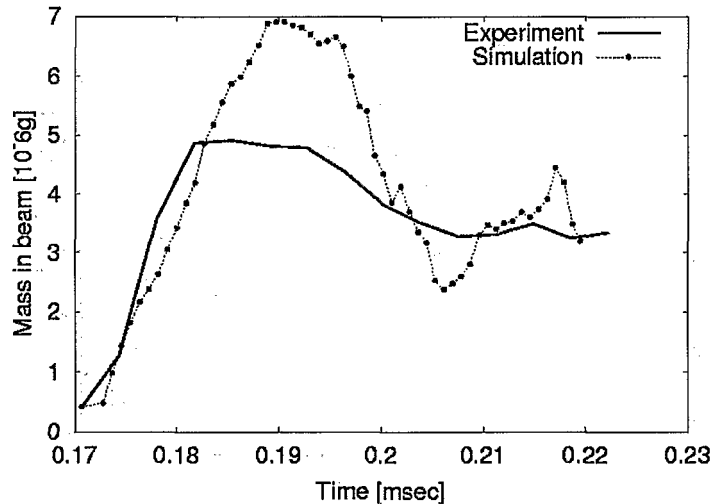


Figure 13. The plot of mass at 1mm from the nozzle exit.

The current work is focused mainly on the simulation of the two-phase mixture resulting from cavitation; the influence of other parameters on spray formation is a subject of further research.

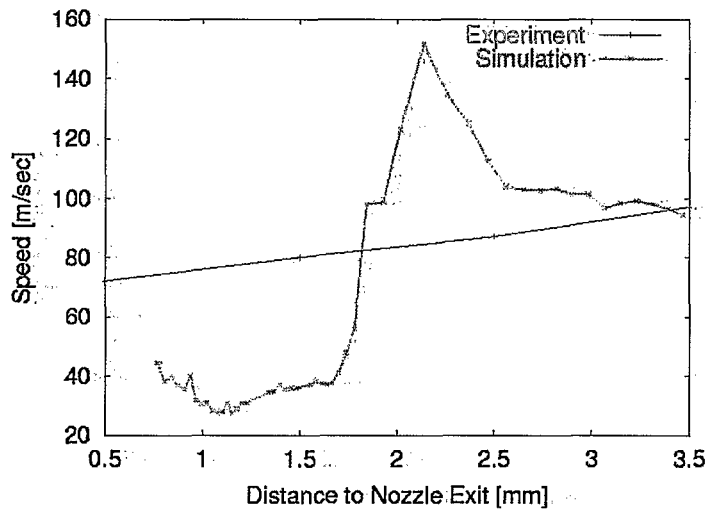


Figure 14: Comparison of jet tip velocity computed from simulations with experimental data.

## References

- [1] MacPhee, A.G., Tate, M.W., and Powell, C.F. X-ray imaging of shock waves generated by high pressure fuel sprays. *Science* **295**: 1261-1263 (2002).
- [2] Poola, R.B. et al. Development of a quantitative measurement of a diesel spray core by using synchrotron X-rays. Eighth International Conference on Liquid Atomization and Spray Systems, Pasadena, CA, July 16-20, 2000.
- [3] Powel, C.F., Yue, Y., Poola, R., and Wang, J. Time resolved measurements of supersonic fuel sprays using synchrotron x-rays. *J. Synchrotron Rad.* **7**: 356-360 (2000).
- [4] Powel, C.F., Yue, Y., Poola, R., Wang, J., Lai, M., and Schaller, J. Quantitative x-ray measurements of a diesel spray core. In *Proc. 14<sup>th</sup> Annual Conference on Liquid Atomization and Spray Systems (ILASS)*, Dearborn, MI, 2001.
- [5] Glimm, J., Kim, M.-N., Li, X.-L., Samulyak, R., and Xu, Z.-L. Jet simulation in a diesel engine. *Proc. Third MIT Conference on Computational Fluid and Solid Mechanics*, 2004, Computational Fluid and Solid Mechanics 2005, ISBN: 0-08-044476-8.
- [6] Xu, Z., Kim, M.-N., Oh, W., Glimm, J., Samulyak, R., Li, X., and Tzanos, C. Atomization of a high speed jet. *J. Multiscale Comp. Eng.* Accepted, 2006.

### 4.3.2 ITAPS: Frontier-Lite

X. Li, Z. Xu, and B. Fix

A general purpose software package for geometry and interface dynamics has been extracted from the FronTier code and is now publicly available.

This package, called FronTier-Lite [1], is designed for users with little training in the algorithms used in the front tracking method and yet deciding to apply the high quality front tracking method to various scientific problems with dynamic front propagation. This code is downloadable from the web, and is accompanied by a web-based testing and evaluation site and extensive web-based documentation.

The software is organized in three levels. The first level deals with the static manifolds and geometry. This includes functions to initialize the interface, optimize an existing interface, and calculate geometry-dependent variables.

The dynamic front package provides a set of functions for the propagation of the interface in a given velocity field and with geometry-dependent velocity functions. The driver function `advance_front()` contains preset pointers to functions based on dimension and algorithms selected by the user. This function will also detect the geometrical and topological correctness of the interface after the propagation. When necessary, it will perform a sequence of operations to guarantee the soundness and quality of the interface before the next propagation step.

The front package allows users to provide their own velocity function for the propagation of the interface. To do so, the user needs to create a data structure containing all necessary parameters for the velocity function. This data structure is cast into a pointer after it is initialized. It will be passed to the velocity function, which is also a pre-assigned anonymous pointer. Inside the velocity function, the pointer of velocity parameters will be cast back to point to the original data structure, and the velocity function retrieves its parameters for the computation of velocity of each marker point.

This front tracking method has been compared with other interface methods on some benchmark problems. In Figure 15, we compare with the level set method. We used the fifth order WENO scheme for the convection of the level set function, while for the front tracking code, we used the fourth order Runge-Kutta method for the point propagation. After 13 revolutions, the fourth order Runge-Kutta method appears to be extremely accurate in the front tracking simulation, while the level set computation begins to show edge smoothing after the second rotation. At the end of the 13<sup>th</sup> circulation, the slot is closed at the top, resulting in a topologically incorrect bifurcation.

Comparison with the volume of fluid method also showed high quality for the front tracking code. One of the benchmark tests is the three-dimensional deformation velocity field described by the velocity functions:

$$u(x, y, z) = 2\sin^2(\pi x) \sin(2\pi y) \sin(2\pi z) \cos(\pi t/T) \quad (1)$$

$$v(x, y, z) = -\sin(2\pi x) \sin^2(\pi y) \sin(2\pi z) \cos(\pi t/T) \quad (2)$$

$$w(x, y, z) = -\sin(2\pi x) \sin(2\pi y) \sin^2(\pi z) \cos(\pi t/T) \quad (3)$$

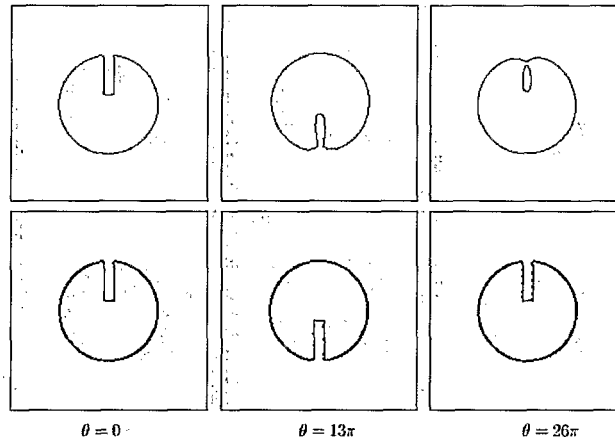


Figure 15. Comparison of slotted disk simulation using high order methods. The upper sequence shows the result of the level set method using the fifth order WENO scheme, and the lower sequence shows the result of front tracking using the fourth order Runge-Kutta method.

The interface evolves dynamically from an initial sphere of radius 0.15 centered at (0.35, 0.35, 0.35) to  $t=1.5$ . The velocity field will then reverse its direction. At  $t=3.0$ , the interface comes back to its initial state. The error comparison with the two PLIC methods is given in Table 1, and shows superior performance for LGB Front Tracking.

Mesh	LGB	Order	CVTNA	Youngs
$32^3$	$5.72 \times 10^{-3}$	3.72	$7.41 \times 10^{-3}$	$7.71 \times 10^{-3}$
$64^3$	$4.33 \times 10^{-4}$	1.82	$1.99 \times 10^{-3}$	$2.78 \times 10^{-3}$
$128^3$	$1.23 \times 10^{-4}$	N/A	$3.09 \times 10^{-4}$	$7.58 \times 10^{-4}$

Table 1:  $L_1$  norms at  $t = 3$  for the LGB method in the three-dimensional deformation simulation compared to the two interface methods used with  $CFL = 0.5$ .

The third level of the front tracking code includes applications to the physical problems, especially the CFD code. The code has recently been used to achieve agreement to experimental value of  $\alpha$ , the mixing rates for the Rayleigh-Taylor instability [2,3]. The code can interoperate with the LLNL code Overture, which provides adaptive mesh refinement (AMR).

The locally-grid based tracking (LGB) provides robust and accurate resolution for the interface geometry. Conservative tracking [4,5,6] has been developed in research code and will be included in a future release. It gives higher accuracy for both the interior and the front.

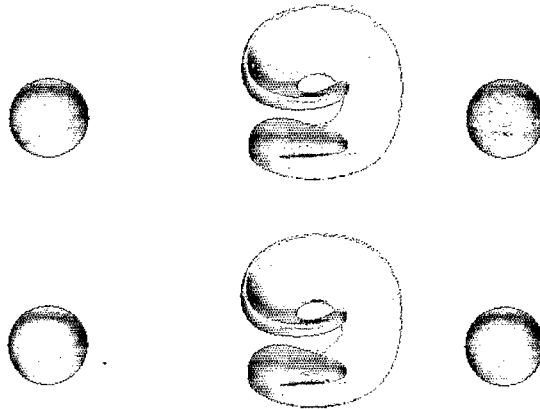


Figure 16: Reversal test of a 3D interface in deformation velocity field with  $CFL = 0.5$ . The sequence above has the mesh of  $64^3$ , and the sequence below has the mesh of  $128^3$ . From left to right are  $t=0, 1.5, 3$  respectively.

## References

- [1] Du, J., Fix, B., Glimm, J., Li, X., Li, Y., and Wu, L. A simple package for front tracking. J. Comp. Phys, **213(2)**: 613-628 (2006). Stony Brook University preprint SUNYSB-AMS-05-02.
- [2] George, E. and Glimm, J. Self similarity of Rayleigh-Taylor mixing rates. Phys. Fluids **7**: 054101-1 – 054101-13 (2005). Stony Brook University preprint SUNYSB-AMS-04-05.
- [3] George, E., Glimm, J., Li, X.L., Li, Y.H., and Liu, X.F. The influence of scale-breaking phenomena on turbulent mixing rates. Phys. Rev. E **73**: 016304-1 – 016304-5, 2006. Stony Brook University preprint SUNYSB-AMS-05-11.
- [4] Glimm, J., Li, X.L., Liu, Y.J., Xu, Z.L., and Zhao, N. Conservative front tracking with improved accuracy. SIAM J. Numerical Analysis **41**: 1926-1947 (2003).

- [5] Liovic, P., Rudman, M., Liow, J.-L., Lakehal, D., and Kothe, D. A 3d unsplit-advection volume tracking algorithm with planarity-preserving interface reconstruction. *Computers and Fluids*. Submitted, 2005.
- [6] Liu, J.-J., Glimm, J., and Li, X.L. A conservative front-tracking method. In *Proceedings of the Tenth International Conference on Hyperbolic Problems: Theory, Numerics, and Applications*, Yokohama Publishers, Osaka, Japan. In press, 2005.

#### 4.3.3 Magnetohydrodynamics of Free Surface Flows

R. Samulyak, J. Du, T. Lu, and J. Glimm

The purpose of this project is to develop novel mathematical models, numerical algorithms, and computational software optimized for modern supercomputers for the numerical simulation of free surface multiphase magnetohydrodynamic (MHD) flows at low magnetic Reynolds numbers, and to perform simulations in support of magnetically confined fusion and advanced accelerator applications. Our primary goal is to gain through mathematical modeling and numerical simulations a better understanding of physics, and improve the design of experiments and devices critical for the DOE mission. The research is being performed in close collaboration with General Atomics, Neutrino Factory/Muon Collider Collaboration, and the SciDAC Center (ITAPS).

Specialized numerical algorithms, optimized for modern computer architectures such as BlueGene, and new models for complex MHD processes in multiphase systems are necessary in order to advance numerical simulations of nuclear fusion devices such as the International Toroidal Experimental Reactor (ITER) and future accelerators. Mathematical models, algorithms, and software being developed under this project enable the simulation of ITER fueling through the injection of small frozen deuterium-tritium pellets, striation instabilities of the pellet ablation channel, liquid hydrogen or lithium jets proposed for the tokamak plasma disruption mitigation, and the simulation of a mercury jet interacting with intense proton pulses in the target system for the Neutrino Factory/Muon Collider.

The system of free surface MHD equations in the low magnetic Reynolds number approximation is a coupled hyperbolic-elliptic system in a geometrically complex moving domain. We have developed a numerical method for solving such equations along with the corresponding parallel software. The numerical method treats the MHD system in an operator split manner. We use the front tracking hydro code FronTier with free interface support for solving the hyperbolic subsystem. The Poisson equation for the electric potential can be solved using techniques for irregular domains described below. FronTier represents interfaces as lower dimensional meshes moving through a volume filling grid. The traditional volume filling finite difference grid supports smooth solutions located in the region between interfaces. The location of the discontinuity and the jump in the solution variables are defined on the lower dimensional grid or interface. The dynamics of the interface comes from the mathematical theory of Riemann solutions, which are idealized solutions of single jump discontinuities for a conservation law. FronTier is capable of tracking 3D interfaces and resolving their topological changes. Some features of the FronTier hyperbolic solvers include the use of high-resolution methods, exact and approximate Riemann solvers, and equation of state models for real materials. The existence of a tracked surface, across which physical parameters and the solution change discontinuously, has important implications for the solution of an elliptic or parabolic system. We have developed an algorithm and software for the dynamic generation of finite element meshes conforming to the interface based on the point shift technique. We recently enhanced capabilities of FronTier-MHD to work with complex interfaces in 3D by developing an elliptic solver based on the embedded boundary method [1]. The method uses a finite volume discretization with control volumes being rectangular grid cells away from

interfaces; and partial cells near the interface (Figure 17). To achieve second order accuracy of the solution, fluxes through irregular cell boundaries are calculated using an interpolation of fluxes in neighboring cells. The corresponding parallel iterative solvers for linear systems of equations are based on PETSc libraries and the algebraic multigrid method implemented in the HyPre software package.

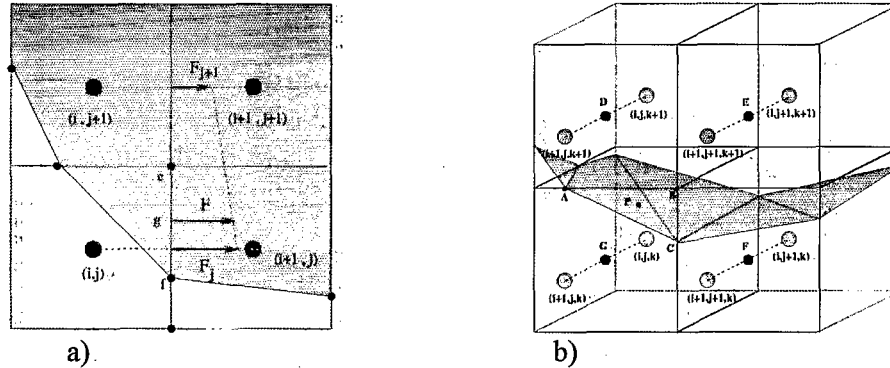


Figure 17: Control volumes near the interface and interpolation of fluxes through irregular cell boundaries in 2D (a) and 3D (b) for the elliptic problem discretization.

An important class of modeling problems related to our targeted applications are the thermodynamic and mechanic properties of materials interacting with intense sources of external energies. These include nonlinear wave phenomena, cavitation and vaporization of fluids, condensation of gases, the ablation of solids by intense laser or electron beams, laser – plasma interaction, and atomic processes (ionization, dissociation and recombination) in gases. Such models have been developed and implemented in the FronTier code. Namely, homo- and heterogeneous (direct numerical simulation) methods for phase transitions such as cavitation, a technique that would allow numerical simulation of the cavitation in liquid hydrogen jet proposed for the mitigation of plasma disruptions, surface ablation models, and models for the interaction of gas/plasma with hot electrons through a simplified solution of kinetic equations have been implemented. Details of the application of the FronTier code to the tokamak fueling and accelerator target problems can be found in [2] and [3], and in Sections 4.3.4 and 4.5.2.

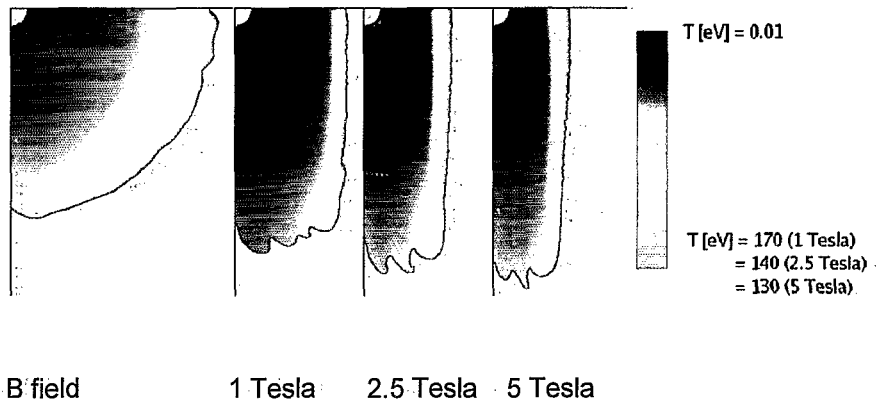


Figure 18: Formation of the pellet ablation channel in tokamak magnetic fields ranging from 0 to 5 Tesla.

## References

[1] Samulyak, R., Du, J., Glimm, J. and Xu, Z.: A numerical algorithm for MHD of free surface flows at low magnetic Reynolds numbers. J. Comp. Phys. Submitted, 2006.

[2] Samulyak, R., Lu, T., and Parks, P. A magnetohydrodynamic simulation of pellet ablation in the electrostatic approximation. Nucl. Fusion. Accepted, 2006

[3] Samulyak, R. and Prykarpatsky, Y. Richtmyer-Meshkov instability in liquid metal flows: influence of cavitation and magnetic fields. Mathematics and Computers in Simulations **65**: 431-446 (2004).

#### 4.3.4 Pellet Ablation in Tokamak Refueling Process

R. Samulyak and T. Lu

The purpose of this research is to develop novel mathematical models, numerical algorithms based on explicit tracking of interfaces, and computational software, and to perform large scale numerical simulations of complex physics processes associated with the injection of frozen deuterium-tritium fuel pellets in tokamaks. This research is being performed in collaboration with General Atomics and Princeton Plasma Physics Laboratory (PPPL).

We have developed a 2D axisymmetric model for the pellet ablation in tokamaks based on the front tracking MHD code [1]. Since the method of front tracking allows the study of multiphysics phenomena in multiphase systems characterized by strong discontinuities in physical properties of the system components, it is ideally suitable to the pellet ablation problem. In our computational model, explicit interfaces separate the solid pellet from the ablated gas, and the cold, dense, and weakly ionized ablation cloud from the highly conducting fusion plasma. Realistic equations of state are employed in different geometrical regions corresponding to different states of matter. The code is capable of simulating the transition of deuterium in the pellet from solid to liquid state under high ablation pressures. A surface ablation model is used for on pellet surface. An electronic heat flux model for the calculation of the thermal energy deposition in the ablation cloud and on the pellet surface uses approximate analytical solutions of kinetic equations. Atomic physics processes in the ablation cloud such as dissociation, recombination, and ionization are taken into account by a plasma equation of state. Using the developed model, we have performed studies of the pellet ablation physics, pellet ablation rate and the lifetime in the magnetic field, and compared with analytical predictions and previous numerical simulations. The 1D version of the model is in excellent agreement with the scaling laws and ablation rate predicted by the theoretical neutral gas shielding model. The 2D axisymmetric simulations agree with previously published numerical results. Our work goes beyond these studies and calculates properties of the ablation channel in magnetic fields and the reduction of the ablation rate.

We showed that a common expectation of the nuclear fusion community regarding the role of the anisotropic heating and magnetic field on the pellet ablation rate was incorrect. The study of striation instabilities, impossible without resolving the detailed physics in the ablation cloud, will be the focus of our future research. These instabilities are not well understood and have a significant impact on the pellet-plasma interaction that will occur in the fueling of burning plasmas in ITER. In the future, we will work with PPPL scientists on the coupling of the M3D code with FrontTier-MHD as a subgrid model for detailed pellet ablation physics, and performing simulations on the BlueGene.

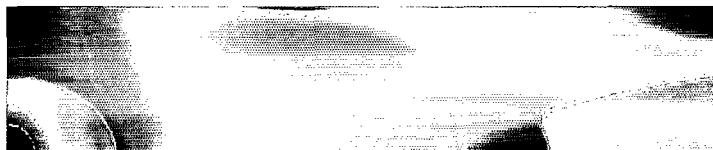


Figure 19. Mach number distribution of the pellet ablation flow demonstrating double transonic layers.

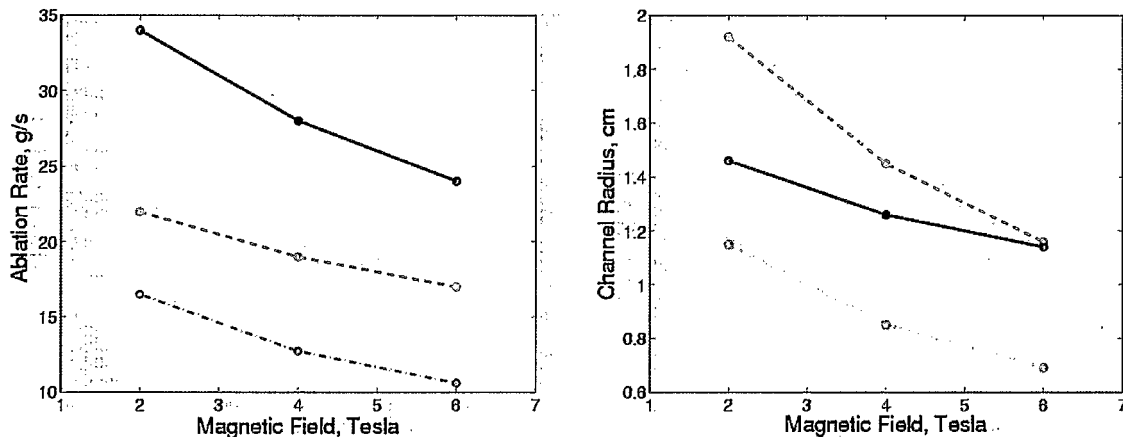


Figure 20: Ablation rate (left) and radius of the ablation channel (right) in tokamak magnetic fields varying from 2 to 6 Tesla at different background electron plasma density  $n_e$  and warm-up time  $t_w$ . Solid line:  $t_w = 10$ ,  $n_e = 1.0e14 \text{ cm}^{-3}$ ; dashed line:  $t_w = 10$ ,  $n_e = 1.6e13 \text{ cm}^{-3}$ ; dotted line:  $t_w = 5$ ,  $n_e = 1.0e14 \text{ cm}^{-3}$ .

#### Reference

[1] Samulyak, R., Lu, T., and Parks, P. A magnetohydrodynamic simulation of pellet ablation in the electrostatic approximation. Nucl. Fusion. Accepted, 2006.

#### 4.3.5 Linear Algebra and QCDOC on Top 500

A. Dubinsky, Y. Gao, C. Jung, R. Bennett, D. Stampf, and Y. Deng

We have ported to the QCDOC supercomputer with up to 8K computing nodes the basic benchmarks from the NPB suite, representing typical Linear Algebra algorithms: BT, LU, SP and MG. LU is a regular-sparse, block (5x5) lower and upper triangular system solver. SP computes the solution of multiple, independent systems of non-diagonally dominant, scalar pentadiagonal equations. BT performs solutions of multiple, independent systems of block tridiagonal equations with a 5x5 block size. MG performs simple multigrid calculations and long-distance communications.

We are examining its strong scalability by using increased numbers of processors to solve fixed-sized problems. We have also investigated weak scaling by studying larger problem sizes. We are testing the impact of the QCDOC 6D-torus interconnect design and distributed memory organization by comparing benchmark results from QCDOC to the results from other supercomputer machines with different structures, such as BlueGene (IBM); and the Beowulf clusters, such as Stony Brook Galaxy computer.

Although QCDOC was designed as a special-purpose machine for Quantum Chromodynamics (QCD) problems, it has shown strong performance and scalability on some applications requiring solution of linear algebra algorithms, and we continue to improve the scalability and increase the system size. An 8K node version has reached 4.61 Tflops at 70.3% efficiency and it would have ranked #109 on the June 2006 Top500 Listing. This performance result has been submitted to the Top500 for the November 2006 Listing and we expect to be listed around #170.



#### 4.3.6 Molecular Dynamics on the QCDOC

P. Rissland, G. Han, G. Martyna, J. Glimm, and Y. Deng

We have designed the first molecular dynamics application package for the QCDOC supercomputer, called MDoC [1-5]. Due to the high complexity and uniqueness of the hardware architecture particularly the network, other molecular simulation packages will require major development to take advantage of QCDOC features. MDoC is custom designed, to fully exploit such unique architectural advances.

Typical bio-molecular systems have thousands, possibly millions, of atoms with timescales at microseconds to milliseconds. Simulating them requires a substantial amount of computational power that can be provided by using supercomputers. The most time consuming calculation for simulating bio-molecules is that of the electrostatic interactions, commonly handled through the use of the Ewald algorithm. We also address other algorithms necessary for molecular simulations, specifically, constant temperature and pressure algorithms using the Melchionna modified Nosé-Hoover method.

We have additionally built an efficient all-gather global operation method for QCDOC [3]. Since QCDOC does not have a network-specific algorithm for global communication like the tree network on BlueGene/L, we designed one specifically for handling the long-range electrostatic interactions. The all-gather method achieves ~60 wire utilization on an eight thousand processor machine.

We have demonstrated the flexibility of QCDOC to run non quantum chromodynamics applications, its primary purpose. Due to the fast communication network and the flexibility of the PowerPC processor, molecular dynamics simulations are capable of 70% parallel efficiency when using 1024 processors on a system with 30,000 particles.

#### References

- [1] Rissland, P. and Deng, Y. Structure and performance of molecular dynamics package MDoC on QCDOC. Parallel Computing. Submitted, 2006.
- [2] Rissland, P. and Deng, Y. Supercomputers and their effects on molecular dynamics. IEEE Potentials, April/May Issue, 9-12 (2005). Chinese Translation appeared in China Computer World Vol. 34 (May 2005).
- [3] Rissland, P. and Deng, Y. All-gather on QCDOC. Parallel Computing. Submitted, 2005.
- [4] Rissland, P. and Deng, Y. Electrostatic force computation on super computers with torus networks. International Conference on Computational Science, Istanbul Turkey, June 2005. Parallel Computing. Submitted, 2006.
- [5] Han, G., Deng, Y., Glimm, J., and Martyna, G. Error and timing analysis of multiping time stepping for MD. Comp. Phys. Comm. August 2006.

#### 4.3.7 Highly Scalable FFT and Molecular Dynamics Algorithms for QCDOC

B. Fang, Y. Deng and G. Martyna

The calculation of long-range electrostatic forces has been one of the biggest challenges in Molecular Dynamics simulation. In parallel computation, it becomes more severe. In the MDoC package, we adopted the basic idea of the Smooth Particle Mesh Ewald (SPME) method and

specifically designed and implemented a scalable scheme for QCDOC [1,2]. There are two aspects in this core that we highlight: the parallel framework for 3D FFT and the employment of a parallel spherical cutoff FFT in the classic SPME method.

In this SPME package, we specifically designed a 3D complex-to-complex FFT and real-to-complex FFT to make full use of the unique six-dimensional torus structure of QCDOC. The actual running time shows that this 3D FFT could scale well up to a 4000 node machine on QCDOC, which is more scalable than the IBM 3D FFT running on BlueGene/L.

The new idea of using a spherical cutoff FFT in the SPME was inspired by the observation that the approximate error for Euler's exponential spline interpolation is not linear, which means that the summation of inaccurate high order Fourier coefficient components is unnecessary. This can save considerable communication time when performing SPME on a parallel machine. The actual running tests show that given a desired tolerant error, the parallel 3D FFT is reduced as a bottleneck if we choose an appropriate interpolation order. Figure 21 shows the parallel efficiency of SPME for the simulation of two biological systems. It turns out that given a tolerant error (a spherical space cutoff  $\hat{g}_c$ ), decreasing the interpolation order  $n$  has more effect than decreasing FFT grid size  $L_{\text{FFT}}$  in increasing the parallel efficiency. However, this is only valid if  $L_{\text{FFT}}$  is of relatively moderate size.

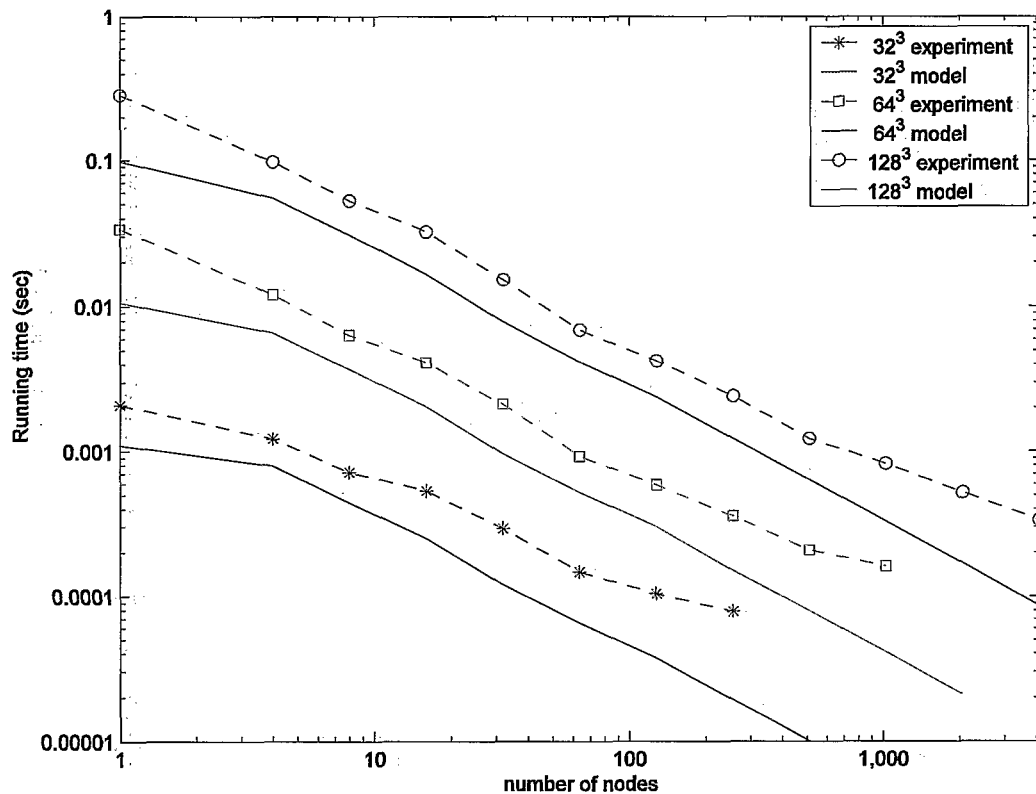


Figure 21. 3D complex-to-complex FFT efficiency on QCDOC. The experiment and model performances, measured by the total running time in solving  $32^3$ ,  $64^3$  and  $128^3$  FFTs on different systems, vs. number of nodes.

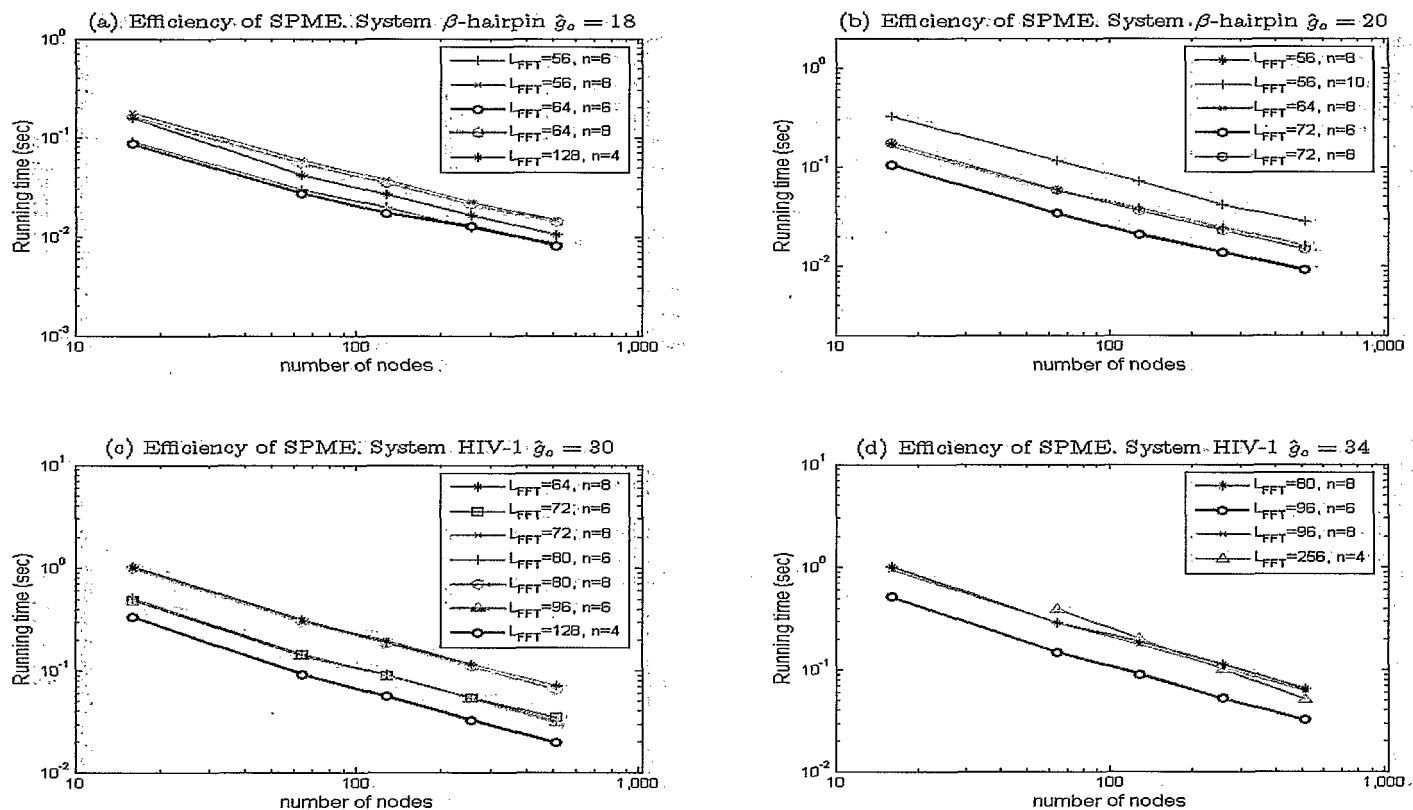


Figure 22: SPME efficiency on QCDOC. Two protein/water systems were simulated: the  $\beta$ -hairpin of Streptococcal Protein G (3579 atoms in total) and HIV-1 Protease (29508 atoms in total). Plots (a) and (b) are for  $\beta$ -hairpin, given two different tolerance errors (spherical cutoffs  $\hat{g}_o$ ); (c) and (d) are for HIV-1. The parameter sets that can provide acceptable approximation errors are plotted. Briefly speaking, spherical cutoff FFT is no longer a bottleneck in this parallel implementation, i.e. if FFT grid size is in relatively moderate size, decreasing interpolation order can provide much better efficiency. However, the condition of the above statement cannot be ignored. For instance, in plot (a) 128 FFT grid size is relatively large and it is the actual bottleneck instead of the interpolation order 4. However, this is not the case in plot (c). The same occurs in plot (d), where the 256 FFT grid size is quite large, so that it is the dominant part of the SPME.

## References

- [1] Fang, B., Deng, Y., and Martyna, G. Performance of 3D FFT on 6D QCDOC torus parallel supercomputers. *Comp. Phys. Comm.* Submitted, 2006.
- [2] Fang, B., Deng, Y., and Martyna, G. Improving the efficiency and accuracy of parallel smooth particle mesh Ewald summation for protein-and-water simulation studies. *Comp. Phys. Comm.* Submitted, 2006.

#### 4.3.8 Electrodynamics on QCDOC and BG/L

R. Bennett and N.L. D'Imperio

We studied the performance of our electrodynamics code MAXSSIM (see 4.4.5) on several massively parallel machines. MAXSSIM is a parallel, scalable, finite-difference time-domain code, which solves Maxwell's equations using the Yee algorithm [1].

It is possible to choose parameters such that communication among processors is a small fraction of the computation time. Thus it provides an important initial test bed for massively parallel machines.

We have ported the code to QCDOC, BlueGene/L, and to the Compaq Alphaserer at the Pittsburgh Supercomputer Center (PSC). We chose a modest sized problem on a  $72 \times 72 \times 2048$  mesh and evaluated the performance on up to 1024 processors.

Figure 23 shows the actual number of floating point operations (sustained performance) as a function of the number of processors on the three machines. Scaling is quite good in all cases (this is a fixed problem size, so this is an example of strong scaling), although the percentage of peak performance achieved on 1024 processors is relatively low – approximately 33.8% on the PSC machine, 22.5% on BG/L and 9.1% on QCDOC.

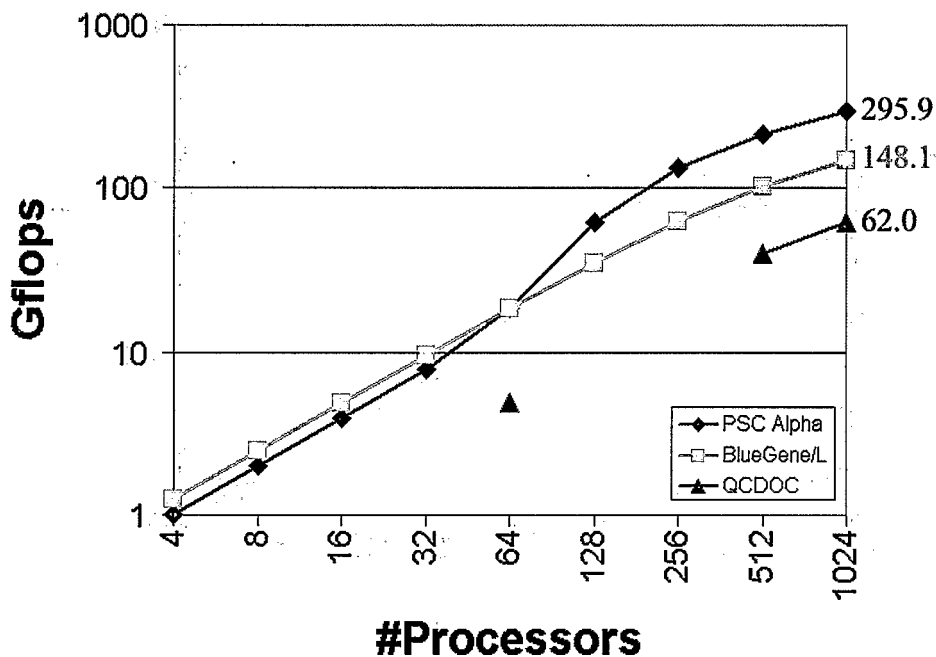


Figure 23.

#### References

- [1] Yee, K.S. Numerical solution of initial boundary value problems in isotropic media. IEEE Trans. Antennas and Propagation **14**: 302 (1996).
- [2] Bennett, R. and D'Imperio, N.L. Performance evaluation of a 3D electrodynamics simulation on massively parallel computers. SIAM Parallel Processing, 2006.

#### 4.3.9 Parallel Heisenberg Spin Model on Supercomputer Architectures

M. McGuigan and R. Bennett

The Heisenberg Spin Model we studied in [1] is used to describe magnetic materials. The model can be extended to a large number of atoms in order to compare with the bulk properties of magnetic materials whose measurement can involve millions or billions of atoms. In the future we will link our study of the Heisenberg model with density functional calculations of its parameters. An important application area is to high-density storage on nanomagnetic materials.

We measured the performance of a parallel Heisenberg spin model using the Monte Carlo method and the Metropolis algorithm on various supercomputer architectures. These architectures include IBM BlueGene/L, PSC Quadrics Cluster, SGI Altix and QCDOC. This assembly of supercomputer systems probes a variety of supercomputer approaches including shared memory, Linux clusters and specialty machines. BlueGene/L, although originally envisioned as a computer for biomolecular simulation, has an efficient implementation of MPI and can be applied to a variety of problems. PSC Quadrics Cluster is a Linux cluster with a quadric interconnect and is also multipurpose. SGI Altix is a shared memory with Numalink interconnect that can also be applied to distributed problems. QCDOC was built originally as a specialized machine for Lattice QCD but has an efficient message passing library called QMP and can also be applied to a wide range of applications, including computational biology [2] and nanoscience.

The Heisenberg spin model of magnetism is defined by the energy

$$E = -J \sum_{nn} \vec{S}_i \cdot \vec{S}_j + \vec{B}_{ext} \cdot \sum_i \vec{S}_i \quad (1)$$

where  $\vec{S}_i$  is a three component spin at lattice site  $i = (i_1, i_2, i_3)$ , the sum over  $nn$  in Eq. (1) is over nearest neighbor lattice sites, and  $J$  is the nearest neighbor coupling. The number of lattice sites or spins along a given direction is given by  $L$ . We parallelized the Heisenberg model by using domain decomposition on large lattices up to 16,777,216 atomic grid points. The number of Monte Carlo steps invoked in the Metropolis algorithm is important in reducing the error in the computation so we wanted to study the number of Monte Carlo steps per second that can be achieved on supercomputer architectures. Previous studies of the Parallel Monte Carlo algorithm for the Ising model were performed in [3] where formulas for the number of Monte Carlo steps per second were obtained.

Two of the main properties of parallel computation we studied were strong scaling and weak scaling. Strong scaling means that we fix the problem size, vary the number of processors, and measure the speedup. This is very important for Monte Carlo simulations because one way of reducing the error in a Monte Carlo simulation is to increase the number of Monte Carlo steps. This can be done in a reasonable time frame by increasing the number of Monte Carlo steps per second. Weak scaling means we vary the problem size and the number of processors such that the execution time is the same. In this way we can obtain higher lattice resolution in the same amount of wall clock time. Results from a weak scaling study of QCDOC are shown in Figure 24. The results are in excellent agreement with weak scaling.

Using the data from the study we were able to fit a Laurent expansion of the form

$$(Steps / s)^{-1} = Time = aL^3 / P + bL^2 / P^{2/3} + c \quad (2)$$

where  $P$  is the number of processors. One way to understand this formula is that the first term represents the time spent in computation and the other two represent time spent in communication

between processors. This form is general enough to include the Ising model parallel performance derived in [3] as well as the Heisenberg model. Note the equation is consistent with weak scaling.

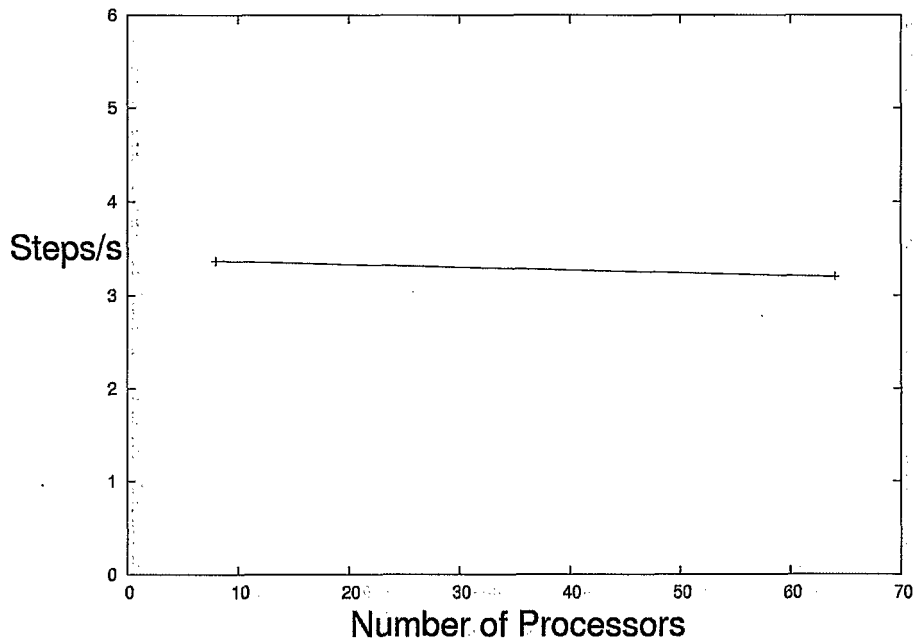


Figure 24. Weak scaling of QCDOC as a function of the number of processors. Performance on the vertical axis is measured in the number of Monte Carlo steps per second.

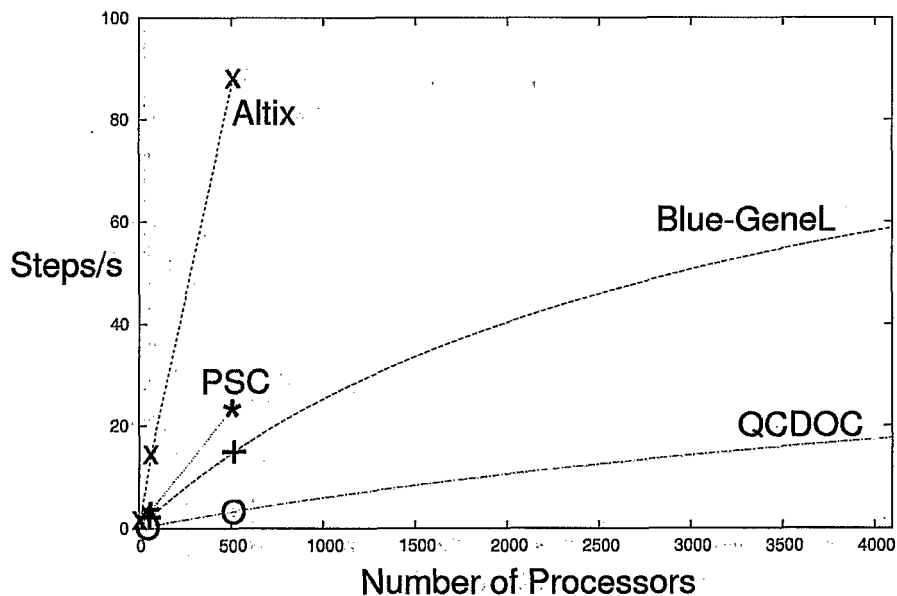


Figure 25. Strong scaling performance curves from various supercomputer architectures using measurements and fits to formula (2).

Fitting the data to the above formula we can obtain an estimate of the performance of the various architectures on large number of processors. The result is shown in Figure 25 for the  $256^3$  lattice and for a variety of computer architectures BlueGene/L, PSC, Quadrics Cluster, SGI Altix and QCDOC applied to the Heisenberg spin model. In all systems we found excellent parallel performance. For example, our results indicate 60 Monte Carlo Steps per second on a  $256^3$  lattice

is possible on a 4096 node BlueGene/L system. This a huge performance gain and will allow the study of large spin systems in the order of a day. On a workstation a similar run could take over a year and make comparison with experiment highly problematic. Our results demonstrate a dramatic increase in productivity in the study of magnetic systems using these leading supercomputer architectures.

## References

- [1] Bennett, R. and McGuigan, M. Parallel Heisenberg spin model performance on supercomputer architectures. ACM 2006 Symposium on Principles and Practice of Parallel Programming, Sept. 2005, submitted.
- [2] Deng, Y., Glimm, J., Davenport, J., Cai, X., and Santos, E. Performance models on QCDOC for molecular dynamics with Coulomb potentials. *Int. J. High Performance Computing Applications* **18(2)**: 183-198 (2004).
- [3] Santos, E. and Muthukrishnan, G. Efficient simulation based on sweep selection for 2D and 3D Ising spin models on hierarchical clusters. *Proceedings of the 18th International Parallel and Distributed Processing Symposium (IPDPS'04)* (2004).

### 4.3.10 Comparison of Two Methods for Elliptic Boundary Value Problems

J. Du, S. Wang, J. Glimm, and R. Samulyak

The purpose of this study is to present a comparison of two popular methods for the solution of elliptic boundary value problems: the embedded boundary method and the mixed finite element methods. The methods are quite different in nature and performance characteristics.

There are many methods for solving the elliptic boundary value/interface problems. Several popular methods have been developed on cartesian meshes for the boundary value/interface problems: the immersed boundary method (IBM) by Perskin [1], the immersed interface method by LeVeque and Li [2], the embedded boundary method by Johansen and Colella [3]. The advantage of these methods is that they are defined on a cartesian grid and there is no need of mesh generation. For the cells away from the boundary/interface, they just use a central finite difference method, which is simple and second order accurate. For the cells that intersect with the boundary/interface, special treatment is needed. Another set of approaches is finite element methods with unstructured mesh, which is more suitable for complex geometries.

We investigate two methods for the elliptic boundary value problem: one is the embedded boundary method (EBM) and the other is the mixed-hybrid finite element method (MFM) [4] with unstructured grid using different basis functions (RT0, BDM1 and RT1). The computational domain and the unstructured mesh are shown in Figure 26. The exponential function is used as the solution of the elliptic problem. Parallel iterative solvers based on PETSc are used for solving the discretized linear system. We compare both the solution errors and performances of the two different methods (Table 1 and Table 2) [5]. Figure 27 shows the L2 error of the solution gradient.

From these results, we draw the following conclusions. Since the embedded boundary method uses a structured cartesian grid, it is relatively easy to implement while it is harder to write the mesh generation program. But after the mesh is given, the discretization is simpler for the mixed finite element method. And it is easier to use the mixed finite element for the elliptic interface problem since the interface is in fact an internal boundary. However, the EBM method must be modified to solve such problems. Also FEM could have higher accuracy if high order basis

function is used. To save computational resources when solving large problems, we could use EBM with automatic mesh refinement since it has better performance for the same order of accuracies.

The EBM has the advantage of fewer unknowns with the same accuracy compared with the MFM. There are two reasons for this. One reason is that the EBM uses a structured grid and the finite volume/central finite difference has super convergence in the mesh. The MFM uses an unstructured grid, and to achieve the same order of accuracy, a higher order basis function space is needed, which means more unknowns. The other reason is that the unknowns for EBM are cell centered and those for the MFM are edge centered. Since the approximate ratio of the vertices to faces to edges is 1:2:3 for a simple large triangle mesh, we know the ratio of the unknowns for the EBM, RT0, BDM1, RT1 is approximately 1:3:6:6. Thus the EBM problem is smaller, which explains why it is much faster.

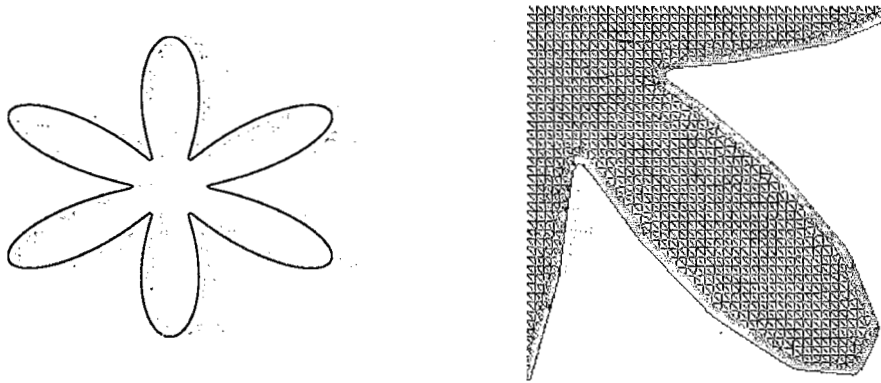


Figure 26. Computational domain and unstructured grid.

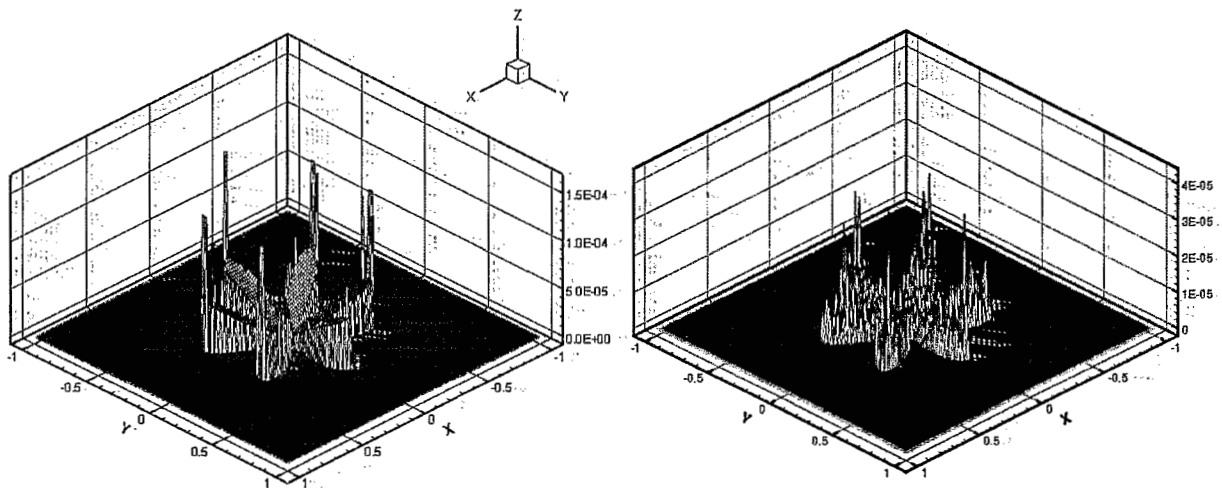


Figure 27: L2 error of solution gradient for EBM (left) and RT1 (right) for 128x128 grid.

Size	EBM	RT0	BDM1	RT1
64 × 64	2.459720e-03	9.367834e-03	7.927028e-04	3.102073e-04
128 × 128	6.567893e-04	6.797467e-03	1.274594e-04	4.007023e-05
256 × 256	1.755489e-04	3.626596e-03	2.486447e-05	1.321007e-05
512 × 512	4.614643e-05	1.754357e-03	6.351849e-06	2.751578e-06

Table 1. Maximum gradient errors on domain boundary.



Mesh Size	EBM				
	error	ratio	time	iterations	unknowns
64 × 64	2.110753e-04	N/A	0.022319	43	1008
128 × 128	5.779287e-05	1.869	0.164115	91	4008
256 × 256	1.472989e-05	1.920	1.438516	209	15738
512 × 512	3.641386e-06	1.952	10.398294	360	61967
Mesh Size	RT0				
	error	ratio	time	iterations	unknowns
64 × 64	1.806532e-03	N/A	0.283352	115	3177
128 × 128	1.142133e-03	0.661	1.624428	218	12341
256 × 256	6.140341e-04	0.895	11.236136	415	47870
512 × 512	3.166839e-04	0.955	79.363582	770	187229
Mesh Size	BDM1				
	error	ratio	time	iterations	unknowns
64 × 64	1.695040e-04	N/A	0.832989	151	6354
128 × 128	5.857866e-05	1.533	5.360611	278	24682
256 × 256	1.641488e-05	1.835	33.627868	461	95740
512 × 512	4.311759e-06	1.929	320.587307	1185	374458
Mesh Size	RT1				
	error	ratio	time	iterations	unknowns
64 × 64	2.203143e-05	N/A	1.212332	151	6354
128 × 128	7.323467e-06	1.589	7.185103	296	24682
256 × 256	2.032626e-06	1.849	45.390103	549	95740
512 × 512	5.312876e-07	1.936	312.086512	1055	374458

Table 2. Convergence and timing comparisons.

## References

- [1] Perskin, C.S. The immersed boundary method. *Acta Numerica* **11**: 479-517 (2002).
- [2] LeVeque, R.J. and Li, Z.L. The immersed interface method for elliptic equations with discontinuous coefficients and singular sources. *SIAM J. Numer. Anal.* **31**: 1019-1044 (1994).
- [3] Johansen, H. and Colella, P. A Cartesian grid embedding boundary method for Poisson's equation on irregular domains. *J. Comput. Phys.* **147**: 60-85 (1998).
- [4] Chavent, G. and Roberts, J.E. A unified physical presentation of mixed, mixed-hybrid finite elements and standard finite difference approximations for the determination of velocities in waterflow problems. *Adv. Water Resources* **14(6)**: 329-348 (1991).
- [5] Du, J., Wang, S., Glimm, J., and Samulyak, R. A comparison study of two methods for elliptic boundary value problems. *SIAM J. Sci. Computing*. Submitted, 2006.

### 4.3.11 Uncertainty Quantification

J. Glimm and Y. Yu.

#### Quantification of Uncertainty and Computer Assisted Decision Making

The need for computer assisted decision making is driven by two related factors. The first is the importance of complex scientific/technical decisions, such as those related to global warming, for which controlled experiments are not feasible. The second is the need for rapid or timely decisions, using incomplete information, such as in shortening the time to market of a product design cycle, mandating a reduction of the role of the human in the loop.

A key issue, and the central one considered here, is an accurate assessment of errors in numerical simulations. Uncertainty quantification (UQ) can be viewed as the process of adding error bars to a simulation prediction. The error bars refer to all sources of uncertainty in the prediction, including data, physics and numerical modeling error. Our approach to uncertainty quantification uses a Bayesian framework. Specifically the Bayesian likelihood is (up to normalization) a probability, which specifies the probability of occurrence of an error of any given size. Our approach is to use solution error models as defining one contribution to this likelihood. We provide a scientific basis for the probabilities associated with numerical solution errors.

We have studied UQ for shock physics simulations [1,2,3], with a focus on statistical analysis of errors in numerical solutions. We decomposed the total simulation error into components attributed to various subproblems, using wave filters to locate significant shock and contact waves. In 1D, a simple error model described the errors introduced at each wave interaction, and (for a spherical geometry) the power law growth or decay of the errors propagated between interactions. A scattering type formula was derived to allow summation of errors propagated through individual interactions to a final error after many interactions [1,2].

The 2D shock interaction problem leads to chaotic interfacial mixing; the central UQ problem here is to define a methodology to describe solution errors for chaotic flow regimes. The challenge is to establish a reliable error analysis for chaotic simulations that do not converge in a pointwise sense, but rather add new complexity with each new level of mesh refinement (see Figure 28). The solution to this conundrum is to look for convergence in averaged quantities, i.e., the statistical moments; and the spatial, temporal, and ensemble averages that will define them. 2D wave filters are used to obtain flow regions having a comparable flow history. Within a single homogeneous flow region, we find that a modest amount of averaging leads to convergent flow quantities under mesh refinement [3].

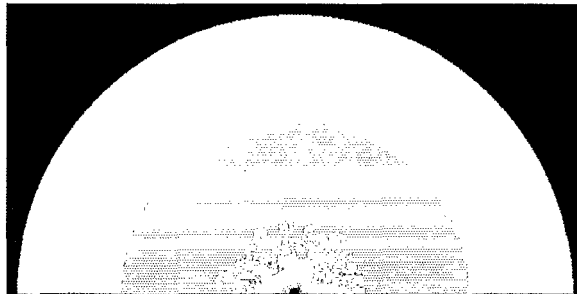


Figure 28. Density plot for a spherical implosion simulation with a perturbed interface. The outer orange-blue boundary is the edge of the computational domain. The red-orange circular boundary is an outgoing reflected shock, and the chaotic inner interface is the object of study. The grid size is 800 x 1600.

## References

- [1] Glimm, J., Grove, J.W., Kang, Y., Lee, T.W., Li, X., Sharp, D.H., Yu, Y., Ye, K., and Zhao, M. Statistical Riemann problems and a composition law for errors in numerical solutions of shock physics problems. *SISC* **26**: 666-697 (2004). University at Stony Brook preprint number SB-AMS-03-11 and LANL report number LA-UR-03-2921.
- [2] Glimm, J., Grove, J.W., Kang, Y., Lee, T., Li, X., Sharp, D.H., Yu, Y., and Zhao, M. Errors in numerical solutions of spherically symmetric shock physics problems. *Contemporary Mathematics* **371**: 173-179 (2005). University at Stony Brook preprint number SB-AMS-04-03 and LANL report number LA-UR-04-0713.

[3] Yu, Y., Zhao, M., Lee, T., Pestieau, M.N., Bo, W., Glimm, J., and Grove, J.W. Uncertainty quantification for chaotic computational fluid dynamics. *J. Comp. Phys.* **217**:200-216, 2006. Stony Brook Preprint number SB-AMS-05-16 and LANL preprint number LA-UR-05-6212.

#### 4.3.12 ViStA: Visual Statistical Analyzer

K. Mueller and W. Zhu

It is well known that once the data become too large in size and/or dimensionality, automated data mining approaches begin to fail. To address these shortcomings, our current efforts target a classify-refine mechanism that inserts the scientist into a tight loop in the data mining process. Interaction with the system is very natural and intuitive, since the user may reorganize, restructure, or fine-tune certain components of the model or portions of the data directly within the visual display.

**Viewing in the native domain.** We have developed a 3D visualization interface that displays correlational data in the native domain. This tool is currently mostly used within the BrainMiner project, where the data and the functional relationships implied by the data are displayed within the brain anatomy. The correlation matrix is a 6-D object and a workable tool for its visualization is a challenge. With user selection of a single row or column, the problem reduces to three-dimensional visualization. For the brain function correlation matrix, these data are presented, along with an MRI volume and a digitized version of the Talairach brain atlas. Both can be sliced in three orthogonal directions and can be overlaid on each other. A basic view with a few regions of interest (ROIs) is shown in Figure 29, which shows the Graphical User Interface (GUI) of our newly developed 3D brain visualization software, along with a basic view of a small number of ROIs embedded into a cut-out area of a normalized/standardized MRI brain. Similar to the 2D viewer, the colors of the ROIs denote the strength of the correlational relationship, on a rainbow scale. The root ROI is colored in yellow. The GUI allows the user to slide the cutting planes up and down and back and forth, to rotate the volume, and to select certain brain surfaces, such as white matter, gray matter, and skull, to be semi-transparently superimposed. The correlation thresholds can also be selected, and many more features are available. The number of ROIs to be displayed, however, can become quite large (about 120-140), which poses challenging problems in the visualization task: in a space too crowded with statistically significant ROIs it becomes very hard, if not impossible, for the user to tell the 3D positions of the individual ROIs. To overcome these difficulties, a number of techniques [1-2] were devised, some of which are illustrated and described in Figure 29.

**Viewing in an abstracted feature domain.** An alternative way to view and edit causal and hierarchical relationships is in a complementary abstracted feature-centric display. We illustrated this type of display in Section 4.2.3 by way of our SpectrumMiner domain application, but similar strategies will also be available to edit causal models for our BrainMiner domain applications. A particular challenge is imposed by time-varying data. Our interactive dendrogram viewer provides the following two mechanisms to cope with this kind of data: (i) the 4D time-slice selector and (ii) the 3D ThemeRiver. We shall describe these two displays in turn.

**The 4D time-slice selector.** The interactive dendrogram represents time-varying data as a cylindrical shape composed of a stack of circular time-slice dendrograms. The time-slice selector shown above the dendrogram in Figure 30 shows the unwrapped outer surface of this cylinder, with each horizontal slice capturing the leaf nodes of one time-slice dendrogram. Patterns in the data distribution over time are clearly visible. Stepping across the time slices will animate the data-related coloring of the dendrogram's arcs.

**The 3D ThemeRiver.** To show the fluctuations of different variables (i.e., nodes in the time-varying dendrogram) over time, we use the ThemeRiver paradigm [3] that visually illustrates multiple variables as parallel streams in which the width of the stream maps to node magnitude. A limitation of the existing ThemeRiver scheme is that only one attribute can be displayed per theme. We therefore have devised a 3D extension, which enables us to display two attributes of each variable in the data stream. The new 3D ThemeRiver [6] can display any ternary relationships within the data, and not just as a function of time.

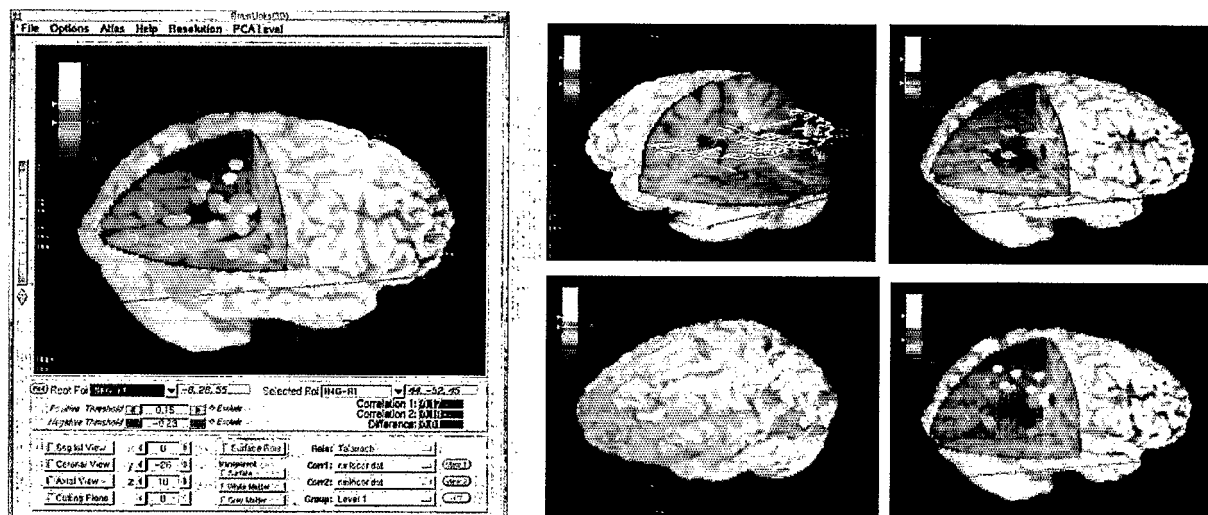


Figure 29. (Left) The present 3D visual interface of BrainMiner, along with a basic view of a small number of ROIs embedded into a cut-out area of a normalized/standardized MRI brain. The display can be switched between three correlation matrices: those of the two studies as well as the difference correlation matrix. The latter display shows the ROIs that have changed, under the influence of the drug, their statistical relationship with respect to the selected root ROI; (center column top:) superimposing a Talairach atlas slice (or an MRI slice) that can be slid up and down the volume; (right column top:) enhancing the ROIs by colored halos or coasters, where the colors code their height and depth on a rainbow color scheme, to aid the perception of 3D depth relations; (center column bottom:) near cortex ROI-correlations projected onto the cortex surface; (right column bottom:) depth cues provided drop lines.

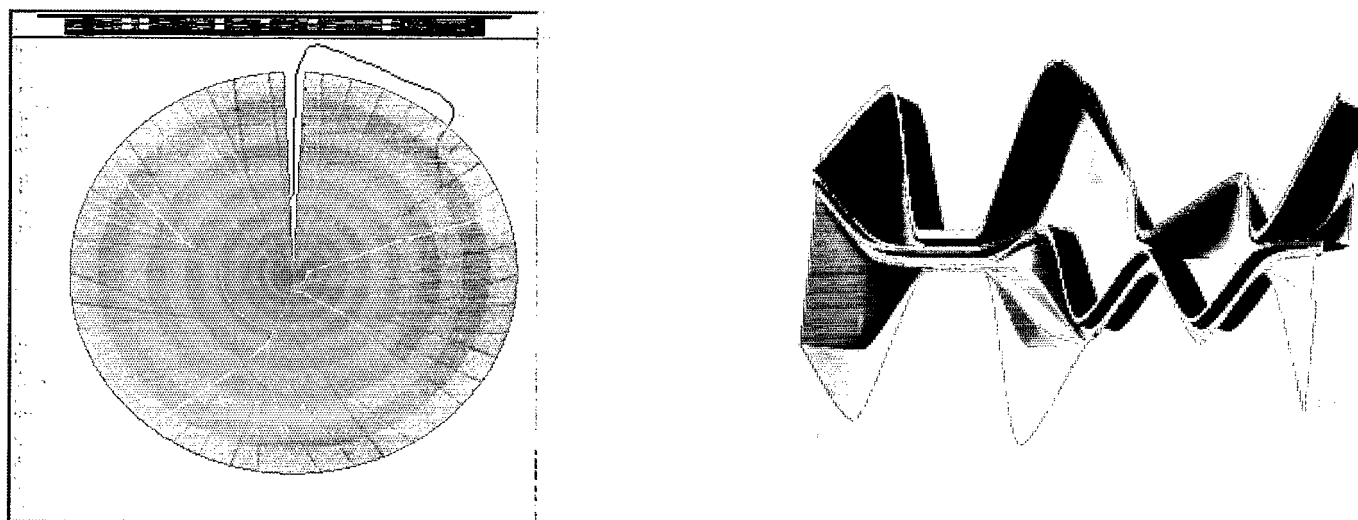


Figure 30. Left: The dendrogram with a user-drawn closed freeform line specifying the time axis-centric surface to be visualized with 3D ThemeRiver. Right: A 3D ThemeRiver visualization of 17 clusters of organic aerosols. Width encodes overall cluster distribution and the height encodes incidence of zinc. The horizontal axis represents the increasing concentrations of ozone in the atmosphere.

## References

- [1] Mueller, K., Welsh, T., Zhu, W., Meade, J., and Volkow, N. Brain Miner: A Visualization Tool for ROI-Based Discovery of Functional Relationships in the Human Brain. New Paradigms in Information Visualization and Manipulation (NPIMV) 2000, Washington DC, November 2000.
- [2] Welsh, T., Mueller, K., Zhu, W., Volkow, N., and Meade, J. Graphical strategies to convey functional relationships in the human brain: A case study. *Visualization '01*, pp. 481-485, San Diego, October 2001.
- [3] Havre, S., Hetzler, E., Whitney, P., and Nowell, L. ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Trans. Visualization and Computer Graphics* **8(1)**: 9-20 (2002).
- [4] Imrich, P., Mugno, R., Mueller, K., Imre, D., Zelenyuk, A., and Zhu, W. Interactive poster: Visual data mining with the interactive dendrogram. *IEEE Information Visualization Symposium'02*, October 2002.
- [5] Volkow, N.D., Zhu, W., Felder, C., Mueller, K., Welsh, T., Wang, G-J., and De Leon, M. Changes in brain functional homogeneity in subjects with Alzheimer's disease. *Psychiatry Research Neuroimaging* **114(1)**: 39-50 (2002).
- [6] Imrich, P., Mueller, K., Imre, D., Zelenyuk, A., and Zhu, W. Interactive Poster: 3D ThemeRiver. *IEEE Information Visualization Symposium '03*, October 2003.
- [7] Imrich, P., Mueller, K., Imre, D., Zelenyuk, A., and Zhu, W. Interactive Poster: A hardware-accelerated rubbersheet focus+context technique for radial dendrograms. *IEEE Information Visualization Symposium'03*, October 2003.
- [8] Zhu, W., Volkow, N.D., Ma, Y., Fowler, J.S., and Wang, G-J. Relationships of ethanol induced changes in brain regional metabolism and motor, behavioral and cognitive functions. *Alcohol and Alcoholism* **39(1)**, 53-58 (2004).
- [9] Kim, J., Zhu, W., Chang, L., Benter, P., and Ernst, T. A unified structural equation modeling approach for the analysis of multi-subject, multivariate functional MRI data. *Human Brain Mapping*. May 22, [Epub ahead of print] (2006).

### 4.3.13 Parallel Visualization of Large Data Sets

R. Bennett, M. McGuigan, G. Smith, J. Spiletic, and S. Tomov

In conjunction with the BNL Medical Department, we are developing a Digital Mouse Brain Atlas [1] that can be used by researchers as a standard reference to map anatomical regions and function associated with imaging studies. For example, in Figure 31 we show a probabilistic Atlas where different colors represent different regions of the mouse brain and darker shades indicate less probable results from an imaging study.

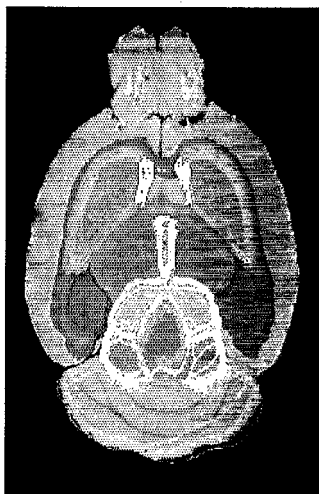


Figure 31. Image from a Probabilistic Mouse Brain Atlas where the colors indicate different anatomical regions and darker shades indicate a lower probability from measurement.



Figure 32. Parallel visualization of the thigh bone of a rat used in osteoporosis studies.

Parallel visualization allows one to handle scientific data sets that are too large or render too slowly on single processor systems. In [2] we implemented interactive parallel visualization by extending popular APIs such as Open Inventor and VTK to support commodity-based clusters. In this work we used Chromium, a popular software tool that provides scalable display technology, to implement our parallel rendering. Figure 32 shows a result from the parallel visualization on four processors of an X-ray tomograph taken of the thigh bone of a rat used in osteoporosis studies. The different colored isosurfaces are handled by separate processors.

The high performance (40Gflops peak) and low cost (\$400) of current graphics cards have given rise to great interest in using Graphics Processing Units (GPUs) for non-graphics scientific computations. In [3] probability-based simulations were implemented on the GPUs and their performance was benchmarked and compared against the CPU with an overall three times speedup for the GPU. Such probability-based simulations have a wide range of applications in physics, biology, chemistry and finance. They are computationally intensive and lend themselves naturally to lattice implementations on GPUs. Specifically we demonstrate the techniques using Monte Carlo simulation of the Ising spin model used in physics to study phase transitions in magnetic materials. The GPU Monte Carlo implementation used Cg (C for graphics) fragment programs, which are executed on the GPU by an OpenGL application. We use a standard "Dynamic texturing" programming model, where the computational domain is modeled by a texture, then a GPU fragment program uses the texture to render an image in an off-screen buffer, and finally the texture is updated from the resulting image. Special care is taken (on the algorithmic and data representation level) to organize the computations in terms of 4D vector operations, which is important in the current GPUs in order to extract maximal performance. Figure 33 shows a direct visualization and computation on the graphics card of the 3D Ising model with red indicating spin up and blue spin down.

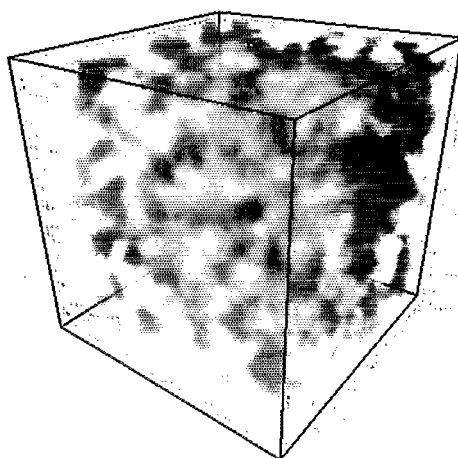


Figure 33. Graphics card computation of 3D Ising model.

#### References

- [1] Benveniste, H., Blackland, S.J., Ma, Y., Hof, P., Bennett, R., McGuigan, M., and Slatest, L. A three-dimensional digital Atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Neuroscience* **135**: 1203-1215 (2005).
- [2] Tomov, S., Bennett, R., McGuigan, M., Peskin, A., Smith, G., and Spiletic, J. Application of interactive parallel visualization for commodity-based clusters using visualization APIs. *Computers & Graphics* **28(2)**: 273-278 (2004).
- [3] Tomov, S., McGuigan, M., Bennett, R., Smith, G., and Spiletic, J. Benchmarking and implementation of probability-based simulations on programmable graphics cards. *Computers & Graphics* **29(1)**: 71-80 (2005).

#### 4.3.14 Visualization Collaboratory

##### **Building a Cross-Institutional Collaboratory for 3D Visualization in Technical Education and Training**

M. McGuigan, R. Bennett, L. Slatest, D. Stampf, BNL, and G. Fong, SUNY Alfred State College

This work builds upon the successes of a previous NSF ATE grant (DUE-0070910) [1]. Therein, two western New York colleges (Alfred State College (ASC) and Jamestown Community College (JCC) partnered with Brookhaven National Laboratory (BNL) in an informal consortium of students, teachers, and scientists focused on the transferal of a cutting-edge, three-dimensional (3D) visualization technology from BNL to the colleges' classrooms. The institutional participants for the current expanded project are the Computational Science Center and the Office of Educational Programs at Brookhaven National Laboratory, SUNY Alfred State College, Bergen Community College, SUNY Jamestown Community College, SUNY Nassau Community College, Rochester Institute of Technology, and SUNY Suffolk County Community College.

The objectives for the current expanded project [2] are: (1) provide students with technical and non-technical experiences not available at their home campuses; (2) update and reinvigorate participating faculty by providing professional development and research collaborations with BNL staff; (3) create 3D Visualization Theaters at the colleges to permit campus-wide involvement in project activities and results; (4) develop specific applications to promote use of the campus

visualization facility within mainstream teaching; (5) provide students with an opportunity to collaboratively design and deliver a course focused on training interested campus clientele in the use of the Visualization Theater; thereby providing both a means to disseminate use of the facility on campus and to enhance students' communication skills and curriculum-specific competencies.

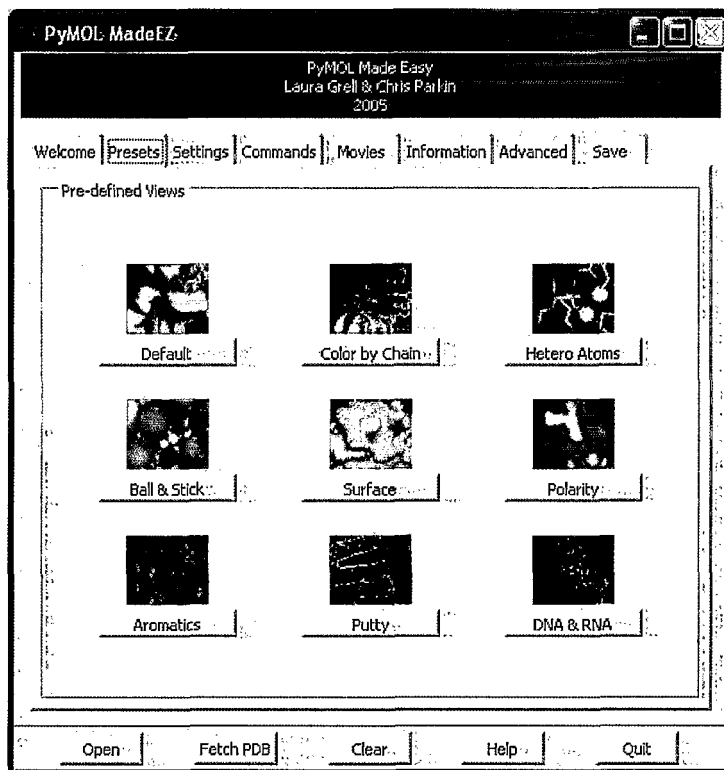
In the first year of this two-year project 18 students from the six campuses worked on nine projects, as developed by the participating faculty members of each college. Two of these projects are discussed here in detail.

### PyMOL Made EZ

L. Grell, C. Parkin, Prof. P. Craig, Rochester Institute of Technology, and L. Slate, BNL

PyMOL [3] is a molecular modeling program that can be used in a wide range of studies within the scientific community. Its ability to produce informative, detailed, stereoscopic images makes it a very powerful tool both in the laboratory and in the classroom. However, the PyMOL user interface is difficult to use, and the need for command line interactions is not user-friendly. Using tools from Python's Tkinter and PMW toolkits, a tabbed graphical user interface (GUI) plugin called PyMOL Made EZ has been created. The GUI contains a series of buttons and entry fields that allow users to easily select given attributes of a molecule and make changes to it with the click of a button.

The PyMOL Made EZ GUI features a notebook style design with separate tabs to distinguish one section from another. Some of its features are: (1) easy to use menus and buttons for the selection and manipulation of molecules; (2) nine predefined molecular views that provide clear and meaningful representations; (3) four movie settings that highlight key molecular aspects; (4) a toggle between PyMOL's normal-viewing and stereoscopic-viewing options; (5) one-click access to primary citation and sequence information; (6) a Chime/PyMOL command converter, and (7) the ability to hide PyMOL's internal interface.



Presets Tab in the PyMOL Made EZ



## Stereoscopic Imaging of Scanning Electron Micrographs

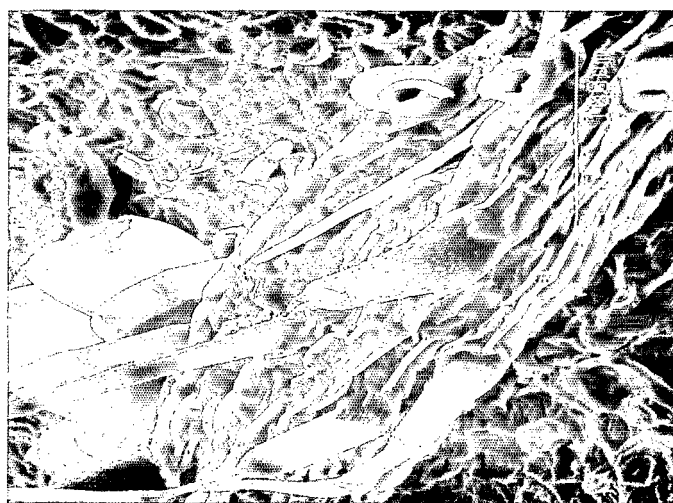
V. Hall, K. Maggio, T. Smith, Prof. S. Beck, Nassau Community College, and J. Spiletic, BNL

Scanning electron micrographs exhibit stereoscopic qualities due to the high depth of field inherent in the use of the scanning electron microscope (SEM). Using the program StereoPhoto Maker [4], selected micrographs, created with the SEM at Nassau Community College [5], are made viewable as stereoscopic images. A routine procedure has been developed that permits movement from trial-and-error to a method that yields consistent results. As in conventional photography, creation of a stereoscopic image using the SEM requires a left-eye and a right-eye image. SEM stage controls permit sample movement in the X and Y axes, 360 degree rotation, and a tilt of 90 degrees toward and 20 degrees away from the signal detector. Images were captured using successive 2.5 degree increments, horizontal and vertical shifting, and sample rotation. Since specimen rotation could not be measured directly, the SEM-generated scale bar was used to measure increments depending upon magnification. The resultant two-dimensional images were opened in StereoPhoto Maker and combined to form a stereoscopic image based on two micrographs of differing perspectives. In addition to the routine method of generating stereo pairs through sample tilting, other methods were also investigated. It was determined that lateral shifting does not allow for the production of a visually satisfying stereoscopic image. However, rotation produced stereoscopic images of equal quality to that of tilted samples. Correlations have been found between surface topography and the initial degree of tilt required.

The specimens shown in the images (Figure 34) possess characteristics that are desirable for stereoscopic viewing, such as physical features of varying surface textures and heights, overlapping structures, and inherent depth. The fish gills specimen creates a quality stereoscopic image because of the overlapping nature of its structures. The structures present in both the foreground and background provide depth to the photo. The leaf is an example of a specimen with varying surface features. Present in the foreground of the photo are hairs and extensive veining. In the background, there are several visible stomates that add variation to the surface topography.



Fish gills



Leaf

Figure 34.

## References

- [1] Fong, G. et al. A Two Year College Cooperative Applied Research Initiative for Faculty and Students in the Engineering and Science Technologies, NSF Proposal DUE 0070910 (1999).

- [2] [http://www.itd.bnl.gov/visualization/nsf\\_ate](http://www.itd.bnl.gov/visualization/nsf_ate)
- [3] <http://pymol.sourceforge.net>
- [4] <http://stereo.jp/eng>
- [5] <http://www.sunynassau.edu/webpages/biology/beck.htm>

#### **4.3.15 Cluster Computing**

N. D'Imperio

Linux clusters have become important providers of computing cycles in the scientific world. (<http://www.ccd.bnl.gov/bcf/cluster/>). The CSC cluster, named Galaxy, consists of 554 Intel P3, P4 and Mac G5 processors running at speeds up to 3.2 GHz. Arranged in dual processor configurations, installed memory is 2 Gbyte per node for the highest performing nodes. A RAID storage array with a 10 terabyte capacity is attached to this machine. Communication is gigabit Ethernet through a Force10 E600 Terascale switch with 240 ports. Parallel computations are performed using the message passing interface, MPI ([www-unix.mcs.anl.gov/mpi](http://www-unix.mcs.anl.gov/mpi)) and through threads using Open MP.

Our goal is to continue to explore the frontier of the commodity component world, as processor, motherboard bus, and memory speeds increase, as fiber cable decreases in price, and as switch speeds increase in performance and decrease in price. We will determine optimal configurations and architectures, within the overall framework of our design. Results from this research will be published and posted on the network, to benefit others seeking a similar price/performance advantage.

The Galaxy computer provides computing cycles to CSC affiliates and projects. Convenient and assured access to local parallel computing is an essential requirement for the development of parallel simulation codes, and it is thus a key component of the CSC strategy to make state-of-the-art computing technology available to benefit BNL scientific programs.

### **4.4 BASIC ENERGY SCIENCES**

The Department of Energy's Office of Basic Energy Sciences supports research in materials and chemical sciences, geoscience, engineering, and energy biosciences. This broad-based program of fundamental research is also responsible for a number of national user facilities, including synchrotron x-ray and neutron sources and electron microscope facilities. CSC has engaged in numerous projects in computational fluid dynamics, many body physics, and optics, which are described here. Most projects are in collaboration with BES scientists and support the BES missions.

#### **4.4.1 Scalable Localizable Density Functional Theory**

K.S. Kang, J.W. Davenport, D. Volja, J. Zheng, D. Keyes, and J. Glimm

Nanoscale science and technology and biology are driving a search for new ways to calculate the electronic properties of clusters containing many thousands of atoms. Density Functional Theory (DFT) has been shown to yield accurate total energies, charge, and spin densities in molecules and crystalline solids but has been limited in the size system that can be treated.

The density functional equations consist of coupled Schrodinger and Poisson equations, which must be solved self consistently. Usually one chooses a basis set (for example Gaussian type orbitals or plane waves), thereby converting the Schrodinger equation into an algebraic eigenvalue problem. The number of such basis functions will scale linearly with  $N$ , the number of atoms in the system. Solving the eigenvalue problem, which is formally dense, will then scale as  $N^3$ . Hence for large systems, there is a premium on efficiency, or reducing the number of basis functions required per atom. Gaussian type orbitals (GTOs) have the advantage of being highly localized in space, decaying like  $\exp(-ar^2)$ . However, they are not maximally efficient, in the sense that many Gaussians are required to accurately represent the wave functions. Plane waves have the disadvantage that they are not localized at all, but rather extend over the whole system. In addition, their use generally requires a “supercell” in which the system is repeated periodically in space, leading to inaccurate treatment of surface and edge effects in some cases.

We solve these problems by using a mixed numerical and localized analytical basis set. We partition the space into nonoverlapping spheres surrounding each atom and an extra-atomic region outside the spheres. Inside the spheres the basis consists of numerical solutions of the Schrodinger (or Dirac) equation for the spherical part of the potential. Outside the spheres, the basis is given by Slater type orbitals, which have the form

$$\phi_{nlm}(\vec{r}) = r^{n-1} \exp(-\zeta r) Y_{lm}(\vec{r})$$

where the  $Y_{lm}$  are spherical harmonics. These functions are not overly localized as are GTOs and achieve the same accuracy more efficiently, typically 5x fewer basis elements per atom.

At each sphere boundary these functions are matched onto the numerical solutions inside the sphere. Such a scheme has already been implemented for periodic systems and is known as LASTO, the Linear Augmented Slater Type Orbital method [1]. It is a local orbital version of the Linear Augmented Plane Wave (LAPW) method, considered the most accurate method for solving the DFT equations.

The Poisson equation is solved on a numerical grid using sparse matrix techniques and the hypre library [2]. Hypre is a set of highly parallel preconditioners and solvers suitable for large sparse systems. The eigenvalue problem is solved using ScaLAPACK [3], a set of codes for the solution of large eigenproblems.

Figure 35 shows the density of states in a 13-atom cobalt-nickel cluster,  $\text{CoNi}_{12}$ . The bond lengths were chosen to be the same as in bulk nickel. The discrete eigenvalues were broadened with a Gaussian.

For the largest systems we use an  $O(N)$  technique known as “divide and conquer” [5]. In this method, a large cluster is partitioned into subsystems, and the charge density is calculated for each. The computational effort scales like  $SN_s^3$ , where  $S$  is the number of subsystems and  $N_s$  is the number of atoms in the subsystem. This scheme is possible because the charge density (more generally, the density matrix) is localized in space even when the eigenfunctions are not [6].

Using divide and conquer, along with the localized basis set provided by STOs, we expect to be able to calculate the charge and spin density of clusters containing up to 5000 atoms.

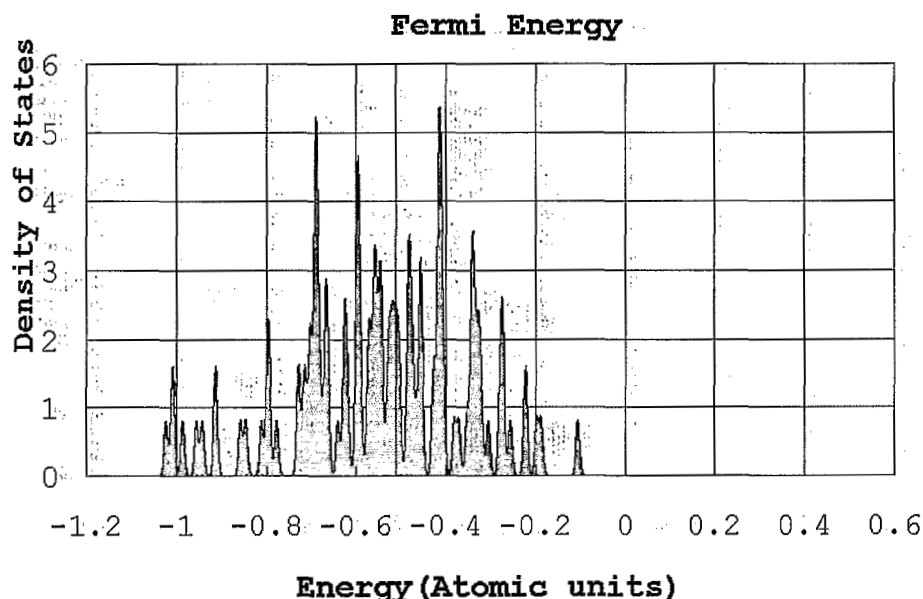


Figure 35: Density of states versus energy for CoNi<sub>12</sub>.

#### References

- [1] Fernando, G.W., Davenport, J.W., Watson, R.E., and Weinert, M. Full-potential linear augmented-Slater type orbital method. *Phys. Rev. B* **40**: 2757 (1989).
- [2] hypre, high performance preconditioners, [www.llnl.gov/CASC/hypre/](http://www.llnl.gov/CASC/hypre/)
- [3] ScaLAPACK, <http://netlib2.cs.utk.edu/scalapack/>
- [4] PETSc, Portable, Extensible Toolkit for Scientific Computing. [www.mcs.anl.gov/petsc/](http://www.mcs.anl.gov/petsc/)
- [5] Yang, W. Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **6**: 1438 (1991).
- [6] Kohn, W. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* **76**: 3168 (1996).

#### 4.4.2 First Principles Calculations of Zinc-Oxide NanoWires

X. Shen, J-C. Zheng, J.W. Davenport, J.T. Muckerman, and P.B. Allen

Nanostructured zinc oxide is a promising material for a new generation of optoelectronic devices [1] and alloys of zinc oxide with gallium nitride have recently been shown to be effective photocatalysts for hydrogen production through water splitting [2]. We have performed first principles calculations of the electronic structure and geometry of ZnO nanowires using a supercell method and the generalized gradient approximation to density functional theory. We used the Quantum Espresso PWscf code [3]. Crystalline ZnO forms in the hexagonal, Wurtzite structure. Our calculations [4] for the lattice constants within the basal plane and along the hexagonal (z) axis differ from experiment by 1.3% and 0.2%, respectively, values which are typical of density functional theory. ZnO wires are naturally formed by cutting out hexagonal shaped fragments with axis along z containing 13 formula

units (26 atoms). Figure 36 compares charge density contours in a plane which contains the z axis for the crystal, and for the unrelaxed and relaxed nanowire. We find that that band gap is increased in the nanowire compared to the bulk crystal while improved photocatalytic properties would require a decrease. This shows the importance of alloying as shown experimentally. Further studies on those systems will be undertaken next.

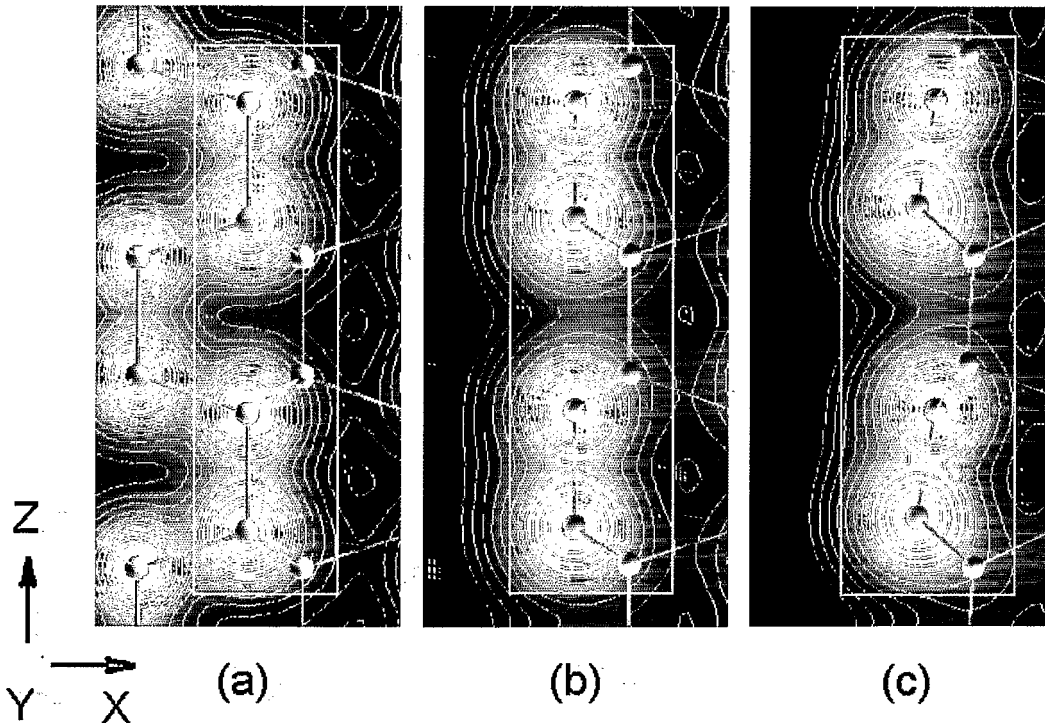


Figure 36: Charge density contours in the x-z plane (z along the hexagonal axis) for (a) bulk ZnO, (b), the unrelaxed nanowire and (c) the relaxed nanowire.

## References

- [1] Wang, Z.L. Zinc oxide nanostructures: growth, properties and applications. *J. Phys.: Condens Matter* **16**: R829-R858 (2004).
- [2] Lu, D., Takata, T., Saito, N., Inoue, Y., and Domen, K. Photocatalyst releasing hydrogen from water. *Nature* **440**: 295 (2006).
- [3] Baroni, S., Dal Corso, A., de Gironcoli, S., and Giannozzi, P. Plane-Wave Self-Consistent Field, <http://www.pwscf.org>.
- [4] Shen, X., Pederson, M.R., Zheng, J-C., Davenport, J.W., Muckerman, J.T., and Allen, P.B. Electronic Structure of ZnO nanowire. *condmat/0610002*.

### 4.4.3 Computational Nanomagnets

M. McGuigan, J. Davenport, and J. Glimm

Nanomagnets are important materials for high-density data storage. Storage densities of  $10^{12}$  bits/in<sup>2</sup> are possible while maintaining thermal stability at room temperature [1]. To design such materials we have devised and validated a computational approach that takes first principles calculations from density functional theory as input to a quantum Heisenberg model using a spin-

wave or magnon description. We established the validity of these methods through comparison to experimental data [2], see Figure 37. We studied an iron nanoribbon of dimension  $14 \times 167 \times 12700$  atoms ( $4\text{nm} \times 48\text{nm} \times 3645\text{nm}$ ) and with a total of 59,385,200 atoms. The comparison shows that computations based on spin wave theory can accurately predict the material properties of nanomagnets below 400K. Above that temperature magnon interactions become important. These higher temperatures to the Curie point can be treated by quantum Monte Carlo methods running on modern supercomputers. The method we developed exceeds by many orders of magnitude the feasible problem size of density functional calculations in terms of the number of atoms treated. The extension of these methods to composite materials is in progress.

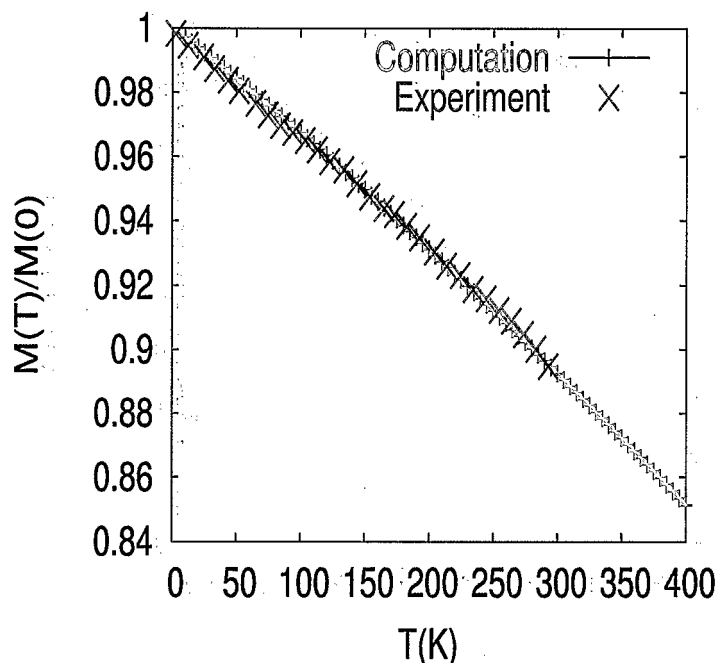


Figure 37: Magnetization curve for an iron nanoribbon with  $14 \times 167 \times 12700$  atoms (total of 59,385,200 atoms). Experimental data from Crespo et al [3].

#### References

- [1] Haast, M. Pattern magnetic thin films for ultra high density recording. Thesis, ISBN-90-3613456 (1999).
- [2] McGuigan, M., Davenport, J., and Glimm, J. Computational approach to finite size and shape effects in iron nanomagnets. Preprint (Oct. 2006).
- [3] Crespo, P., Gonzalez, J., Hernando, A., and Yndurain, F. Spin-wave excitations in ribbon-shaped Fe nanoparticles. Phys. Rev. B **69**: 012403 (2004).

#### 4.4.4 DFT Study of Isocyanide Adsorption on Gold (111) Surface.

Y. Gilman, P.B. Allen, and M. Hybertsen

Isocyanides are molecules terminated with the  $-\text{NC}$  group, which can serve as the "alligator clip" to connect a molecule to metallic electrodes. Isocyanides may find application in molecular electronics. Several groups successfully fabricated self-assembled monolayers of di-isocyanide

molecules sandwiched between two gold electrodes and studied electrical transport through them [1]. At Stony Brook University, work is under way to fabricate a molecular transistor based on a single di-isocyanide molecule.

We study the details of adsorption of two simple isocyanides, HNC and  $\text{CH}_3\text{NC}$  on the gold (111) surface by density functional theory (DFT) methods. The standard slab calculation scheme is employed, in which the real surface is modeled by the surface of a slab of several atomic layers of gold (see Figure 38). Monolayers at various coverages are considered: one molecule per one Au surface unit cell, one per three, and one per four. We use the WIEN2K [2] DFT code based on the full-potential linearized augmented plane wave (FPLAPW) method. The generalized-gradient approximation (GGA) of Perdew et al. (PBE) is chosen for the exchange-correlation potential.

We find that molecules can adsorb only at the site on top of a Au atom, which agrees with experimental evidence [3,4]. The adsorption is weak with the adsorption energy of only 0.2 eV for both HNC and  $\text{CH}_3\text{NC}$  molecules. When the local density approximation (LDA) is chosen for the exchange-correlation potential, the results are qualitatively different: adsorption is possible at all sites, with hollow site preferred, and the adsorption energy is about 1 eV. In the absence of adsorption energy measurements, it is difficult to tell whether GGA description of these systems is accurate. We repeated the calculations for the CO molecule (isoelectronic to CNH) for which experimental values of adsorption energies on Au are known. We find that while the LDA overestimates adsorption energies, GGA underestimates them by about 0.2 eV.

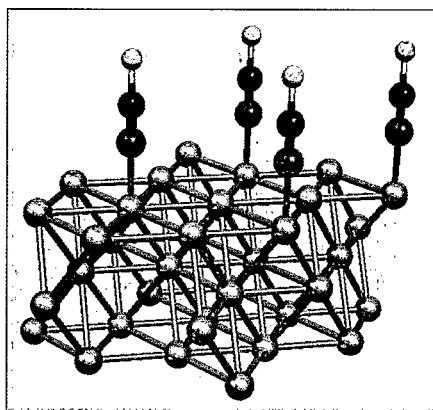


Figure 38.

## References

- [1] Chen, J., Calvet, L.C., Reed, M.A., Carr, D.W., Grubisha, D.S., and Bennett, D.W. *Chem. Phys. Lett.* **313**: 741 (1999); Lee J.-O., Lientschnig, G., Wiertz, F., Struijk, M., Janssen, R.A.J., Egberink, R., Reinhoudt, D.N., Hadley, P., and Dekker, C. *Nano Letters* **3**: 113 (2003).
- [2] Blaha, P., Schwarz, K., Madsen, G.K.H., Kvasnicka, D., and Luitz, J. *WIEN2k, An Augmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties* (Karlheinz Schwarz, Techn. Universit.at Wien, Austria), 2001. ISBN 3-9501031-1-2.
- [3] Robertson, M.J. and Angelici, R.J. *Langmuir* **10**: 1488 (1994).
- [4] Gilman, Y., Allen, P.B., and Hybertsen, M.S. density-functional study of adsorption of isocyanides on the gold (111) surface. *Phys. Rev. B*. Submitted, 2006.

#### 4.4.5 Direct Numerical Simulation of Multiphase Flows with Phase Transitions

T. Lu, R. Samulyak, Z. Xu, and J. Glimm

An accurate description of cavitation and wave propagation in cavitating and bubbly fluids is a key problem in modeling and simulation of hydrodynamic processes in a variety of applications ranging from marine engineering to high-energy physics. We are interested in the study of cavitation and bubbly fluids occurring in high speed liquid jets such as diesel jets in fuel injectors and liquid mercury targets that interact with high intensity proton pulses. Such targets are key components of advanced accelerators such as the Spallation Neutron Source (<http://www.sns.gov>) and Muon Collider/Neutrino Factory ([www.cap.bnl.gov/mumu](http://www.cap.bnl.gov/mumu)). Another important application is the simulation of cavitation in a high-speed liquid lithium or helium jet in a device proposed for the plasma disruption mitigation in tokamaks.

Modeling of the cavitation is a complex multiscale and multiphysics problem involving the description of thermodynamic properties of liquids under strong external fields and nonlinear wave phenomena in a multiphase system. In the direct numerical simulation of bubbly flows, a liquid-vapor or liquid-non-dissolvable-gas mixture is represented as a system of one-phase domains (vapor bubbles, for instance) separated by free interfaces. The FronTier code is capable of tracking simultaneously a large number of interfaces and resolving their topological changes (the breakup and merger of droplets) in two- and three-dimensional spaces. Though computationally intensive, such an approach is potentially very accurate in treating important effects in bubbly flows including bubble oscillations, heat transfer, drag, viscosity, and surface tension. The method makes it possible to resolve spatial scales smaller than the typical distance between bubbles, and to model some non-equilibrium thermodynamic features such as finite critical tension in cavitating liquids. The direct method has been validated through the comparison of numerical simulations with theoretical predictions and classical experiments on linear (sound) and nonlinear (shock) waves in bubbly fluids [1-3] (see Figure 39) and applied to the study of the Muon Collider and Spallation Neutron Source targets (see Sections 4.5.1 and 4.5.4).

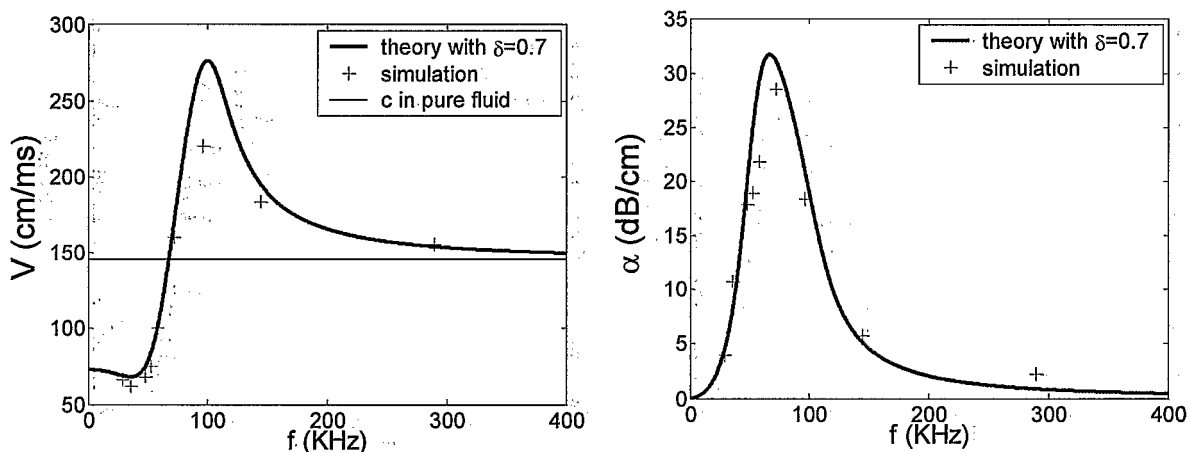


Figure 39. Phase velocity and attenuation rate of linear waves in bubbly water.

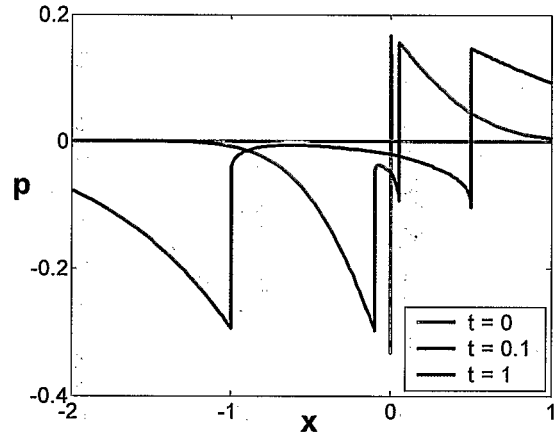
Numerical simulation of the cavitation presents an additional level of complexity compared to the simulation of wave phenomena in fluids containing small, non-dissolvable gas bubbles. The problem is associated with the dynamic creation and collapse of bubbles in the computational domain. The corresponding numerical models and software routines have been developed and implemented in the FronTier code. The phase transition rate is proportional to the deviation of the vapor pressure from the saturated pressure.



$$M_{ev} = \frac{\alpha}{\sqrt{2\pi RT}} (p_{sat}(T) - p_{vap}),$$

where  $\alpha$  is the condensation coefficient. The Riemann problem for phase transition has been studied using the method of viscous profiles. The transient waves in realistic (thermal conductive) phase transitions with Riemann data have been demonstrated, in particular, by the analytical solution for the linear waves, as shown in Figure 40.

Figure 40. Analytical solution for the transient linear pressure wave induced by the temperature difference in the Riemann data.



To account for the phase transition induced mass transfer across the liquid-vapor interface; a numerical scheme has been developed in the frame of front tracking. A non-local Riemann solver governing the evolution of interfaces has been implemented and a numerical technique has been introduced to account for the thin thermal layer. The algorithm has been validated and applied to physical problems such as the jet breakup in diesel engines [2].

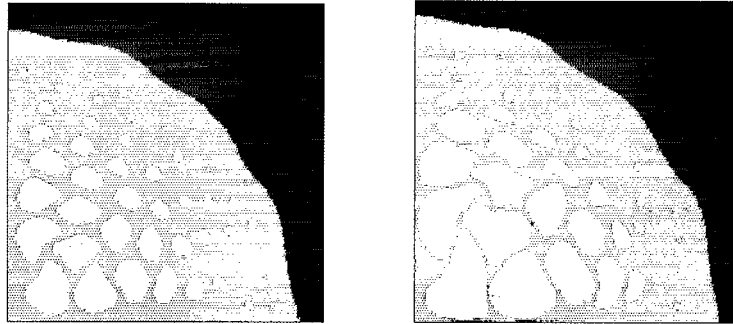


Figure 41. 2D numerical simulation of cavitation in the mercury jet after the interaction with a proton pulse depositing 100 J/g of energy into the jet. Density distribution on the jet cross-section is shown: red is mercury, light blue is the rarefied gas in cavitation bubbles, and dark blue is the ambient gas. Time frames are 20 (left) and 40 (right) microseconds.

#### References

- [1] Lu, T., Samulyak, R., Prykarpatsky, Y., Glimm, J., Xu, Z., and Kim, M.N. Comparison of heterogeneous and homogenized numerical models of cavitation. *Int. J. Multiscale Comp. Eng.* **4(3)** (2006).
- [2] Xu, Z., Kim, M., Oh, W., Glimm, J., Samulyak, R., Li, X., Lu, T., and Tzanos, C. Atomization of a high speed jet. *Int. J. Multiscale Comp. Eng.* Accepted, 2006.

[3] Xu, Z., Lu, T., Samulyak, R., Glimm, J., and Ji, X.M. Dynamic phase boundaries for compressible fluids. SIAM J. Computing. Submitted, 2006.

## **4.5 HIGH ENERGY AND NUCLEAR PHYSICS**

New major physics facilities are central to the vitality of the nation's HENP programs. Due to their expense and complexity, simulation codes are integral to the design process. Design errors have been a major contributor to past instances of project cancellations and missed target performance. For this reason, design teams will increasingly use high performance computing, developed in national, multi-laboratory collaborations to address simulation issues in the design, operation, and control of accelerators.

### **4.5.1 Neutrino Factory/Muon Collider Target**

R. Samulyak, J. Glimm, and J. Du

In order to understand the fundamental structure of matter and energy, an advance in the energy frontier of particle accelerators is required. Advances in high-energy particle physics are paced by advances in accelerator facilities. The aim of the multi-institutional research group Neutrino Factory/Muon Collider Collaboration is to explore the feasibility of a high-energy, high-luminosity muon-muon collider and a neutrino factory. For more information, visit the Neutrino Factory/Muon Collider Collaboration home page (<http://www.cap.bnl.gov/mumu>). However, several challenging technological problems remain to be solved in the collider design in order to achieve potential advantages of greatly increased particle energies over traditional electron-positron machines (linear colliders). One of the most important problems is to create an effective target able to generate high-flux muon beams. The need to operate high atomic number material targets in particle accelerators that will be able to withstand intense thermal shock has led to the exploration of free liquid jets as potential target candidates for the proposed Muon Collider. The target will contain a series of mercury jet pulses of about 1 cm in diameter and 30 cm in length. Each pulse will be shot at a velocity of 30 m/s into a 20 Tesla magnetic field at a small angle to the axis of the magnetic field. When the jet reaches the center of the magnet it will be hit with a 2 ns proton pulse. Every proton pulse will deposit about 100 J/g of energy in the mercury.

Numerical simulations of hydro and MHD processes in the target can reduce the amount of costly experiments and help to optimize target parameters. Such simulations present a challenging problem of computational science. They require mathematical modeling of complex flows undergoing phase transitions (cavitation) and numerical methods for solving MHD equations in complex geometries. Our numerical studies have been performed using FrontTier-MHD, a magnetohydrodynamic extension of the front tracking hydro code FrontTier (for more details, see Section 4.3.3). Numerical simulations of the Neutrino Factory/Muon Collider target have already achieved important results. In our previous works, we addressed problems of the interaction of mercury jet with proton pulses, the evolution of Richtmyer-Meshkov instabilities on the jet surface and the jet breakup, stabilizing effect of the magnetic field, and cavitation on mercury caused by strong rarefaction waves. Simulations agreed with experiments carried out at BNL's Alternating Gradient Synchrotron and at CERN.

We have demonstrated that at the design parameters, the distortion of the jet entering a non-uniform magnetic field is significant, and that reduces the effective cross-section of the interaction with the proton pulse (Figure 43). The study has led to the change of design parameters of the future experiment MERIT. Current work is focused on large-scale 3D MHD simulations of the mercury jet interaction with proton pulses.

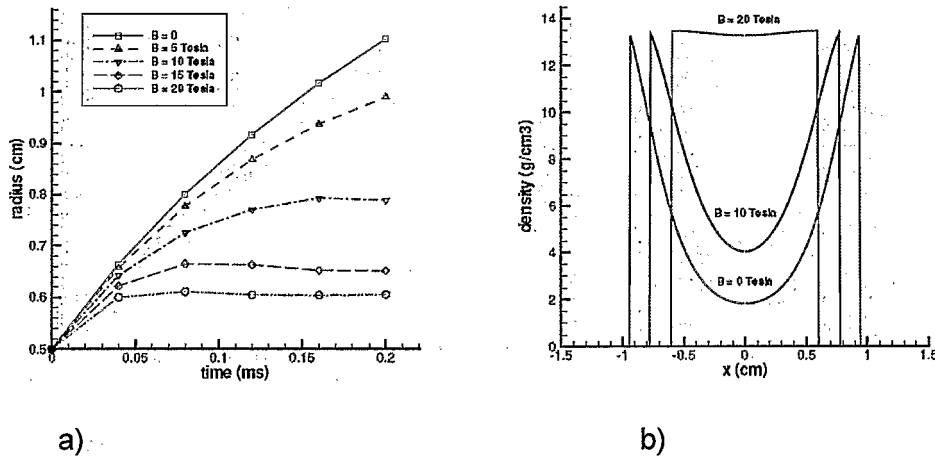


Figure 42: Mercury jet evolution in magnetic fields ranging from 0 to 20 Tesla. a) Evolution of the mercury jet radius; b) Average density profile in the jet cross-section at 130 microseconds.

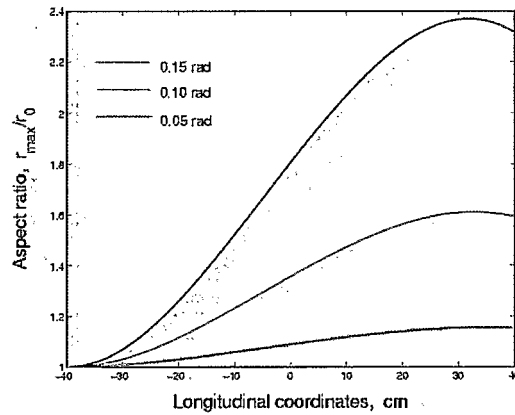


Figure 43: Aspect ratio of the cross-section of the mercury jet entering a 15 Tesla solenoid at three different angles with respect to the solenoid axis. Zero of the longitudinal coordinate corresponds to the center of the 1 meter long solenoid. The jet nozzle is placed inside the solenoid at  $x = -40$  cm in coordinate system.

## References

- [1] Lu, R., Samulyak, R., Prykarpatsky, Y., Glimm, J., Xu, Z., and Kim, M.N. Comparison of heterogeneous and homogenized numerical models of cavitation. *Int. J. Multiscale Comp. Eng.* **4(3)** (2006).
- [2] Samulyak, R., Du, J., Glimm, J., and Xu, Z. A numerical algorithm for MHD of free surface flows at low magnetic Reynolds numbers. *J. Comp. Phys.* Submitted, 2006.
- [3] Samulyak, R., and Prykarpatsky, Y. Richtmyer-Meshkov instability in liquid metal flows: influence of cavitation and magnetic fields. *Mathematics and Computers in Simulations* **65**: 431–446 (2004).

## 4.5.2 Modeling of Wake Fields and Impedances in Accelerators

R. Samulyak

The electromagnetic interaction of an intensive charged particle beam with its vacuum chamber surroundings in an accelerator plays an important role for the beam dynamics and collective beam

instabilities. Wake fields, generated by a moving particle in the accelerator pipe and objects such as RF cavities, bellows, stripline monitors, etc., affect the motion of particles in the tail part of the beam causing parasitic loss, beam energy spread, and instabilities. The effect of wake fields is usually of the same order of magnitude as the space charge effect. While the space charge forces approach zero in the ultrarelativistic limit, wake fields remain finite for an ultrarelativistic beam due to resistivity of the accelerator walls and non-smoothness of the chamber (existence of RF cavities, bellows etc.). The effect of wake fields is an increasingly important issue since operating regimes are continually moving towards higher currents and smaller bunches. To avoid collective beam instabilities that limit accelerator performance, an accurate numerical modeling of wake fields and their interaction with the beam is necessary.

In the traditional approach for including wake field forces in an accelerator code, the total impedance budget for the accelerator ring is calculated or experimentally measured and the corresponding forces are applied to tracked particles once per beam turn. Such a cumulative force approach is not sufficient for the simulation of beam instabilities caused by wake fields. It is less accurate than the 3D computation of the space charge that has already been developed in advanced accelerator modeling codes, including the MaryLie/Impact and Synergia codes. We have developed a model that accounts for the fine structure of particle beams and distributes wake fields in the accelerator chamber. The corresponding theoretical model is based on the expansion of the particle beam in terms of the multipole moments and the notion of the wake function, which allows elimination of the complex temporal behavior of the electromagnetic field between the incident charge creating the wake field and the test charge. The wake function describes the response of the accelerator chamber element to a delta-functional, pulse carrying,  $m$ -th moment. Wake functions are independent of beam properties and are defined totally by properties of the accelerator chamber.

The wake field algorithm is coded as a parallel Fortran 90 module which performs charge deposition of macroparticle beams on a 3D grid, expansion of the corresponding charge distribution into multipole moments, computation of wake functions and wake field forces, and interpolation of the wake field forces from the grid to macroparticles. The module has been implemented in the MaryLie/Impact and Synergia codes. In the current numerical implementation, most of the accelerator chamber elements (resistive pipe, RF cavity, etc.) have associated analytical wake field models valid under certain approximations. Analytical wake field models are beneficial for the study of long-range wake fields and their multiturn effect on the beam dynamics in circular accelerators. To study wake field effects in accelerator elements that cannot be accurately approximated by analytical models, wake functions in a tabular format can also be used. The corresponding data can be obtained through accurate numerical solutions of the full Maxwell system of equations using commercial (MAFIA) or public domain electromagnetic codes.

We have recently developed and implemented a subgrid model for the calculation of resistive wake fields on the sub-millimeter length scale. The resolution of such short-range wake fields is very important for the dynamics and energy balance of accelerator beams. The subgrid model was validated using analytical theory of wake fields for Gaussian bunches. We have shown that for a short Gaussian bunch, particles located at  $0.5\sigma$  ahead of the bunch center lose energy due to wake forces, and particles located at  $1.8\sigma$  behind the bunch center gain energy (Figure 44), opposite to the space charge effect. Such calculation was not possible without the subgrid model, since sub-millimeter range wakes in the vicinity of a particle inducing wake fields are responsible for the energy loss of trailing particles.

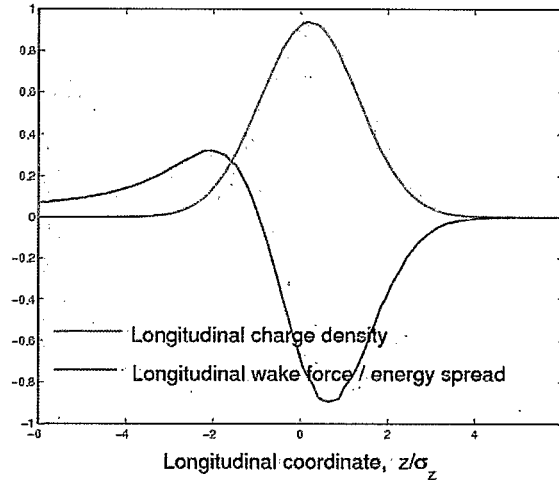


Figure 44: Normalized longitudinal resistive wake field force of a short Gaussian bunch calculated using the wake field code with a subgrid model for sub-millimeter scale wake fields.

#### Reference

[1] Ryne, R. et al. SciDAC advances and applications in computational beam dynamics. J. Physics: Conf. Series **16**: 210-214 (2005).

#### 4.5.3 Unified Accelerator Library: SIMBAD

N. D'Imperio and A. Luccio

Working in collaboration with computational physicists at BNL's Collider Accelerator Department, we have continued development on a Particle-in-Cell (PIC) [1] code to model collective beam effects in two and three dimensions. The code was formerly known as ORBIT and is being integrated into the Unified Accelerator Libraries (UAL) [2] as the SIMBAD component for the modeling of space charge.

UAL provides a framework in which the tracking of particles takes place by pushing a “herd” of macroparticles through a lattice using the TEAPOT code, which has also been integrated in UAL. Space charge is calculated separately using SIMBAD. Once all particles have reached a certain location, their charge density is calculated by binning to a grid. The potential  $\Phi$  is found by solving the Poisson equation with the perfectly conducting wall boundary conditions:

$$\Delta\Phi = -\rho/\epsilon, \quad \rho_{\text{wall}}(x,y,z) = 0.$$

Space charge force components (with coefficients to account for both the electrostatic and magnetic action) are calculated as derivatives of the potential and applied to each macro particle in the transverse direction. Longitudinal space charge is calculated by binning the particles longitudinally and following the formalism presented in [1]. In a ring with long longitudinal bunches, the transverse motion can be uncoupled from the longitudinal, and the Poisson equation can be solved in parallel in many longitudinal beam segments.

The parallelization of SIMBAD is implemented differently in 2-d as opposed to 3-d. In 2-d, the parallelization is implemented by dividing the macro particles among the processes and collectively calculating the forces on each mesh. In 3-d, each process takes a number of longitudinal slices and calculates the space charge forces only within its own slices. This parallelization requires load

balancing that considers both the number of particles and the number of slices. This is accomplished using a genetic algorithm [3, 4]. Figure 45 shows parallel decomposition of the bunch and the resulting division of the beam. Figure 46 shows the parallel efficiency of a simulation using the load balancing algorithm [4].

The code was used to model high intensity beams in the Alternating Gradient Synchrotron at BNL to study its suitability as a proton driver [3]. In addition, the code is being used to study the long-term impedance effects of the SIS100 accelerator currently being designed and built at GSI in Darmstadt, Germany.

UAL has been ported to run on the BlueGene Light supercomputer where beam-beam interaction will be simulated.

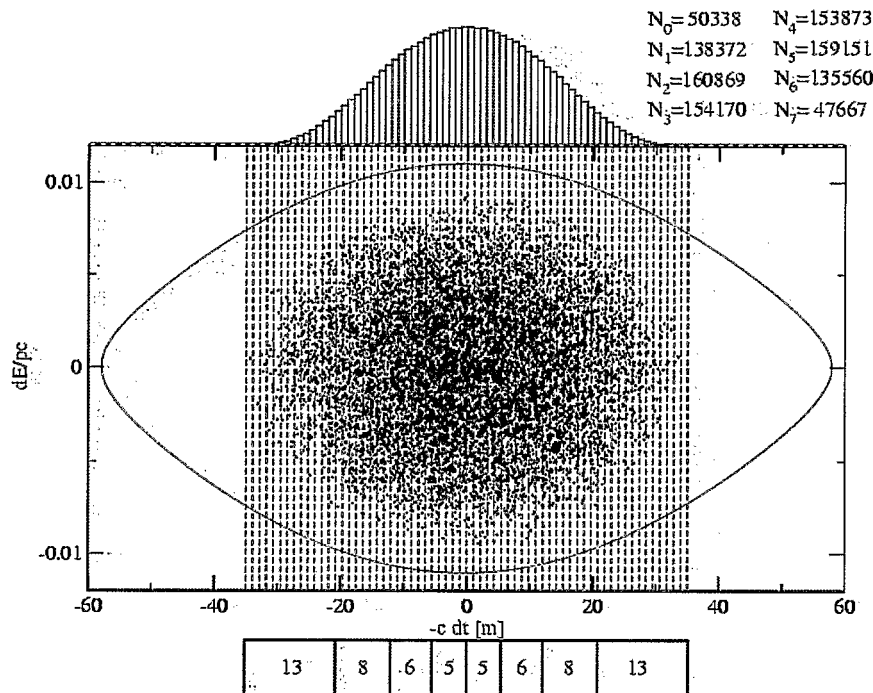


Figure 45. Bunch decomposition and load balancing. (a) A bunch of macro particles in an RF bucket with accompanying lines to indicate the boundaries for each of 8 processes. (b) Genetic algorithm produces an optimal balance between number of particles per process and number of slices per process.

### Parallel Efficiency

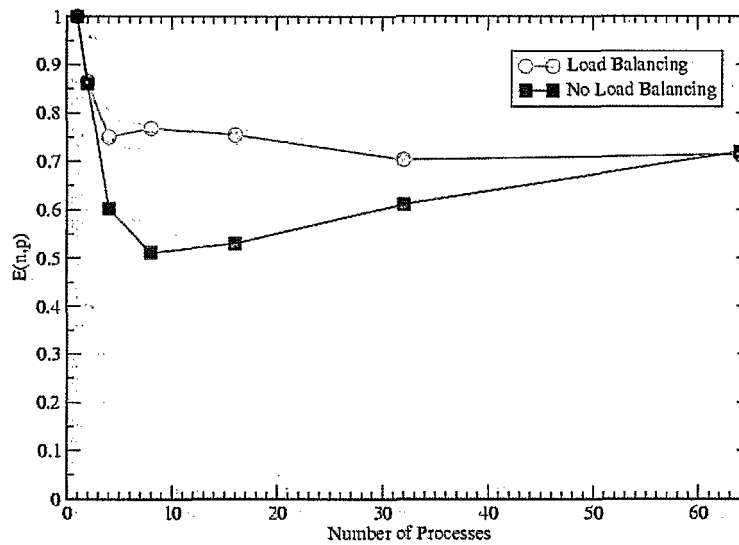


Figure 46. Efficiency improvements of up to 50% are realized with load balancing.

$$E(n,p) = \frac{S(n,p)}{P}, \quad S(n,p) = \frac{T_s(n)}{T_p(n,p)}, \quad T_s(n)$$

Is the runtime of a serial solution with problem size n and

$$T_p(n,p)$$

Is the runtime of a parallel solution with P processes.

### References

- [1] Chao, A.W. *Physics of Collective Beam Instabilities in High Energy Accelerators*. Wiley, 1993.
- [2] Malitsky, N. and Talman, R. Unified Accelerator Libraries. In *Computational Accelerator Physics*, AIP Conference Proceedings 391, J.J. Bisognano and A.A. Mondelli, Eds., pp. 337-342 (1997).
- [3] Luccio, A. and D'Imperio, N. Simulation of the AGS as a proton driver. ICFA Beam Dynamics Mini Workshop on Space Charge Simulation, Oxford, UK, Trinity College, April 2-4, 2003, [www.bd.fnal/icfa/workshops](http://www.bd.fnal/icfa/workshops) (2003).
- [4] D'Imperio, N.L., et al. Parallel 3-D Space Charge Calculations in the Unified Accelerator Library. EPAC 2006, July 2006, <http://accelconf.web.cern.ch/AccelConf/e06/INDEX.htm>.

#### 4.5.4 Spallation Neutron Source

R. Samulyak and T. Lu

The Spallation Neutron Source (SNS) is an accelerator-based neutron source being built in Oak Ridge, Tennessee, by the U.S. Department of Energy (<http://www.sns.gov>). The SNS will provide the most intense pulsed neutron beams in the world for scientific research and industrial development.

The proposed liquid mercury target design for the Spallation Neutron Source (see Figure 47) includes a main flow region inside a stainless steel structure where mercury enters from the sides, flows around a baffle into the proton beam path, and exits out the center. A cooling jacket that wraps from bottom to top around the target is used to cool the target window through which the proton beam enters. The stainless steel target structure is approximately  $0.5 \times 0.4 \times 0.15 \text{ m}^3$ .

One of the most important issues associated with using liquid metals as targets for pulsed proton beams is withstanding the loads caused by the rapid pressure increase resulting from the intense heating of the liquid metal from a single pulse of protons. This heating occurs essentially instantaneously, compared to acoustic time scales; therefore, the mercury undergoes a large pressure increase. In addition to a set of difficult engineering problems associated, for instance, with the design of windows able to withstand large thermal gradients and shocks, recent experiments with an SNS target prototype uncovered yet another problem critical to the target lifetime. They showed pitting of stainless steel surfaces that were in contact with mercury subject to large pressure pulses induced by the collapse of cavitation bubbles [1]. Due to the cavitation-induced erosion, it will be necessary to replace the target after two weeks of operation at frequency 60 Hz of a 1 MW proton pulse. To extend the target lifetime, future research efforts will be concentrated in two areas, each of which should lead to reduction of the erosion damage:

- Evaluation of cavitation-resistant materials and coatings.
- Investigation of mitigation techniques such as introduction of non-dissolvable bubbles into the system.

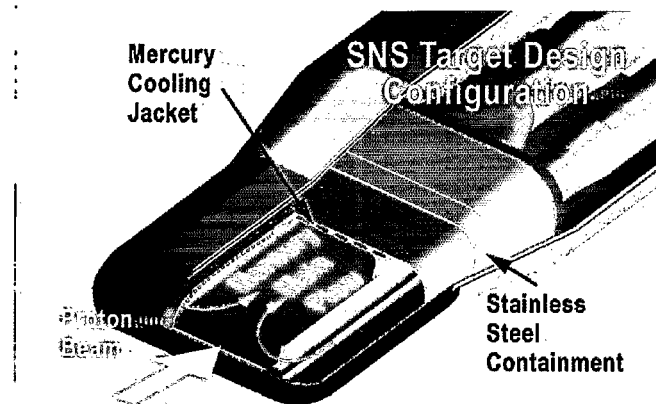


Figure 47. SNS mercury target design.

We have applied the direct numerical simulation technique for bubbly fluids (see Section 4.4.3) to the study of pressure mitigation through the injection of non-dissolvable gas bubbles near the target front window. We have found that while the bubbly layer indeed causes a significant reduction of pressure during 200 microseconds, large transient pressure oscillations exist for a short period of time ( $< 100$  microseconds) after the proton beam energy deposition (see Figure 48). We have studied the formation and evolution of cavitation bubbles in mercury caused by the pressure distributions depicted in Figure 48. The collapse pressure of cavitation bubbles was calculated by solving the Keller equation. The mitigation efficiency, estimated by performing statistical averages of pressure peaks, was found to be dependent on the parameters of the bubbly layer such as the volume fraction and average bubble size. For example, a bubbly layer with the average bubble size  $R = 0.5 \text{ mm}$  and a 0.53% volume fraction reduces the integral effect of cavitation-induced pressure peaks by 50 times.



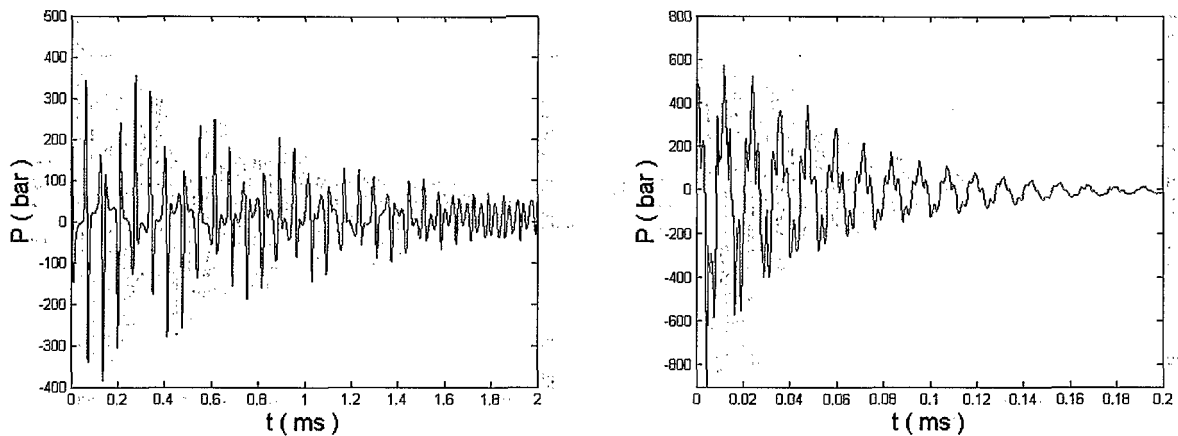


Figure 48. Proton pulse induced pressure peaks on the entrance window in the pure mercury (left) and mercury containing gas bubbles (right).

#### References

- [1] Status Report on Mercury Target Related Issues, SNS-101060100-TR0006-R00, July 2002.
- [2] Lu, T., Samulyak, R., and Glimm, J. Direct numerical simulation of bubbly flows and application to cavitation mitigation. J. Fluid Eng. In press, 2006.

### 4.5.5 Grid Computing: MonALISA

#### Development and Use of MonALISA High Level Monitoring Services for the STAR Unified Meta-Scheduler

E. Efstathiadis, L. Hajdu, J. Lauret, and I. Legrand (CalTech)

As a Particle Physics Data Grid (PPDG) cross team-project we study, develop, implement and evaluate a set of tools that allow Meta-Schedulers to take advantage of a consistent set of shared information (such as information needed for complex decision making mechanisms) across both local and Grid Resource Management Systems. We demonstrate the usefulness of such tools within the MonALISA monitoring framework and the STAR Unified Meta-Scheduler.

We define the requirements and schema by which one can consistently provide queue attributes for the most common batch systems and evaluate the best scalable and lightweight approach to access the monitored parameters from a client perspective and, in particular, the feasibility of accessing real-time and aggregate information. Client programs are envisioned to function in a non-centralized, fault tolerant fashion. We believe that such developments could highly benefit Grid laboratory efforts such as the Grid3+ and the Open Science Grid (OSG).

The MonALISA (Monitoring Agents in A Large Integrated Services Architecture) system provides a distributed monitoring service. It is based on a scalable Dynamic Distributed Services Architecture (DDSA) that is implemented using JINI/JAVA and WSDL/SOAP technologies. The scalability of the system derives from the use of autonomous, multi-threaded station servers to host a variety of loosely coupled, self-describing, dynamic services; the ability of each service to register itself and then to be discovered and used by other services or clients that require such information; and the ability of all services and clients subscribing to a set of events (state changes) in the system to be notified automatically. The framework integrates several existing monitoring tools and procedures to collect parameters describing computational nodes, applications and network performance. It has built-in SNMP support and network-performance monitoring algorithms that enable it to monitor end-to-end network performance, as well as the performance and state of site facilities in a Grid.

The core of the MonALISA monitoring service is based on a multithreaded system (the monitoring service) used to perform the many data collection tasks in parallel, independently. It is designed to easily integrate existing monitoring tools and procedures and to provide this information in a dynamic, self-describing way to any other services or clients. MonALISA services are organized in groups and their group attribute is used for registration and discovery. Each service registers with a set of JINI Lookup Discovery Service (LUS), as a member of a group, and having a set of attributes. The LUSs are also JINI services and may be registered with other LUSs, resulting in a distributed and reliable network for registration of services. Services also provide the code base for the proxies that other services or clients will need to instantiate for using it.

A generic framework for building pseudo-clients for the MonALISA services was developed. This has been used for creating dedicated web service repositories with selected information from specific groups of monitoring services. The pseudo-clients use the same LUSs approach to find all the active MonALISA services from a specified set of groups and subscribe to these services with a list of predicates and filters. These predicates or filters specify the information the pseudo-client wants to collect from all the services. Pseudo-clients store received values from the running services in a local MySQL database. A Tomcat based servlet engine is used to provide a flexible way to present global data and to construct on the fly graphical charts for current or customized historical values, on demand. Multiple Web Repositories can easily be created to globally describe the services running in a distributed environment.

Queue monitoring data collected using custom MonALISA modules at each site are cached locally using pseudo-clients with fail-over capabilities. This way the monitoring data from all services that have joined a group are available to policies implemented into the Meta-Scheduler that choose the appropriate queue for the submitted job.

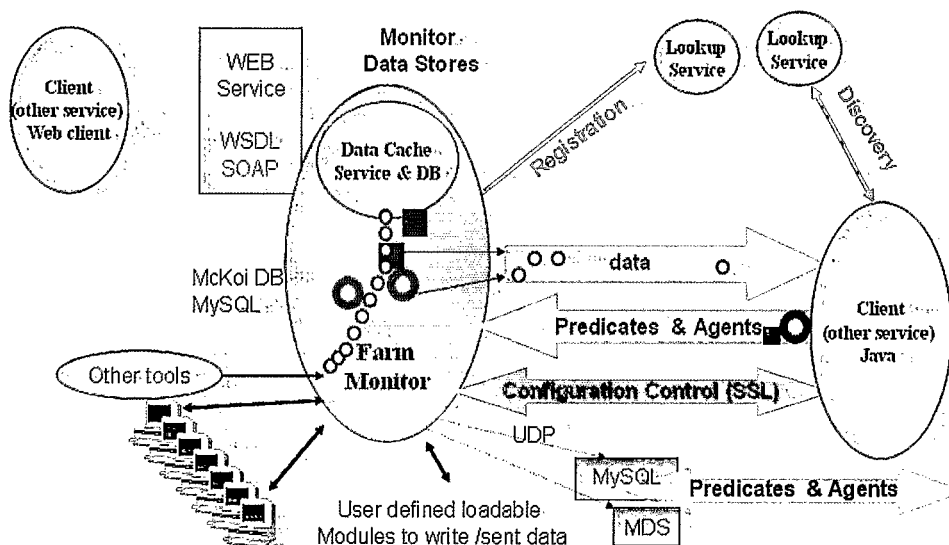


Figure 49. The MonALISA monitoring service.

In addition, a MonALISA service has been deployed at BNL for the monitoring needs of the UltraLight project [4]. UltraLight is a collaboration of experimental physicists and network engineers whose purpose is to provide the network advances required to enable petabyte-scale analysis of globally distributed data. Current Grid-based infrastructures provide massive computing and storage resources, but are currently limited by their treatment of the network as an external, passive, and largely unmanaged resource. The goals of UltraLight are to:

- Develop and deploy prototype global services that broaden existing Grid computing systems by promoting the network as an actively managed component.
- Integrate and test UltraLight in Grid-based physics production and analysis systems currently under development in ATLAS and CMS.
- Engineer and operate a trans- and intercontinental optical network testbed, including high-speed data caches and computing clusters, with U.S. nodes in California, Illinois, Florida, Michigan and Massachusetts, and overseas nodes in Europe, Asia and South America.

MonALISA can be used to monitor and control network devices, such as routers and photonic switches. Since it gathers information system-wide, MonALISA is able to generate global views of the prevailing network connectivity, to identify network or end-system problems and act on them strategically, or locally as required. Services that take decisions based on these (global) system views can be created and deployed: for example, mobile agents that are able to provide optimized dynamic routing for distributed applications have recently been added to MonALISA.

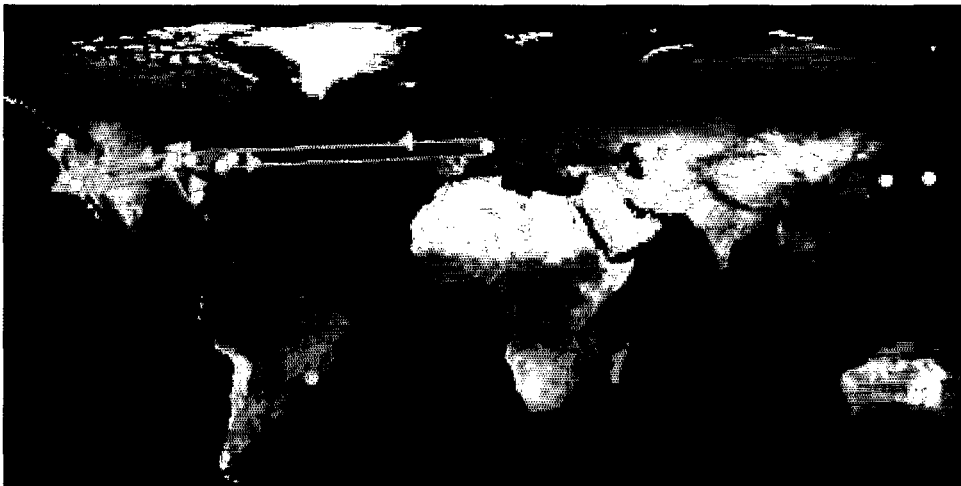


Figure 50: The UltraLight testbed includes sites in the Americas, Europe and Asia.

## References

- [1] Efstathiadis, E. et al. Development and use of MonALISA high-level monitoring services for meta-schedulers. Computing in High Energy and Nuclear Physics (CHEP '04), Interlaken, Switzerland, Sept. 27 - Oct. 1, 2004.
- [2] [http://www.ivdgl.org/documents/document\\_server/uploaded\\_documents/doc--998--ivdgl\\_star\\_monitoring.ppt](http://www.ivdgl.org/documents/document_server/uploaded_documents/doc--998--ivdgl_star_monitoring.ppt)
- [3] <http://indico.cern.ch/getFile.py/access?contribId=393&sessionId=7&resId=0&materialId=slides&confId=0>
- [4] <http://ultralight.caltech.edu/web-site/ultralight/html/index.html>.
- [5] Efstathiadis, E., et al. Status of the QCDOC project at BNL. International Workshop on QCDOC and BlueGene, Edinburgh, UK, October 4-6, 2005.
- [6] Bennett, G.W., et al. Final report of the E821 muon anomalous magnetic moment measurement at BNL. Phys. Rev. D. **73** (2006).

